

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

«До захисту допущено»
В.О. Завідувача кафедри
_____ О.Л. Тимощук

Дипломна робота

**на здобуття ступеня бакалавра
за освітньо-професійною програмою «Системний аналіз і управління»
спеціальності 124 «Системний аналіз»
на тему: «Аналіз та прогнозування кредитних ризиків у банківській
сфері»**

Виконала:

студентка IV курсу, групи КА-64
Садома Юлія Володимирівна _____

Керівник:

доцент, д.т.н., доцент кафедри ММСА,
Кузнєцова Наталія Володимирівна _____

Консультант з економічного розділу:

доцент, к.е.н., доцент кафедри ТТРЕ
Шевчук Олена Анатоліївна _____

Консультант з нормоконтролю:

доцент, к.т.н., доцент кафедри ММСА
Коваленко Анатолій Єпіфанович _____

Рецензент:

доцент, к.т.н., доцент кафедри СП
Безносик Олександр Юрійович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студентка _____

Київ – 2020 року

**Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»**

Інститут прикладного системного аналізу

Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

В.о.завідувача кафедри

_____ Оксана ТИМОЩУК

«__» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

Садомі Юлії Володимирівні

1. Тема роботи «Аналіз та прогнозування кредитних ризиків у банківській сфері», керівник роботи Кузнєцова Наталія Володимирівна, д.т.н., доцент, затверджені наказом по університету від «25» травня 2020 р .№ 1143-с

2. Термін подання студентом роботи 08 травня 2020 року _____

3. Вихідні дані до роботи

1. Фінансові показники АТ КБ «ПРИВАТБАНК» з 2001-го року по 2019-й рік.
2. Мова програмування Python
3. Середовище розробки – PyCharm Community Edition 2018.3.5
4. Бібліотеки та модулі, що використовувалися: NumPy, Pandas, StatsModels, Sklearn, Matplotlib, Math, Missingpy

4. Зміст роботи

1. Проаналізувати існуючі методи, що використовуються для прогнозування даних, фільтрації даних та заповнення пропусків
 2. Обрати методи для прогнозування даних, фільтрації даних та заповнення пропусків.
 3. Розробити програмний продукт на базі обраних математичних методів;
 4. Застосувати розроблений програмний продукт на реальних даних, проаналізувати результати
 5. Виконати економічний аналіз програмного продукту
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

1. Презентація

6. Консультанти розділів роботи*

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	к.е.н., доц. Шевчук О. А.	21.04.20	29.05.20

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	15.04.20	
2	Аналіз існуючих методів для прогнозування даних, фільтрації даних та заповнення пропусків	24.04.20	

* Якщо визначені консультанти. Консультантом не може бути зазначено керівника дипломної роботи.

3	Пошук даних та формування вибірки	7.05.20	
4	Розробка програмного продукту	12.05.20	
5	Тестування продукту на реальних даних, аналіз результатів	25.05.20	
6	Оформлення дипломної роботи	29.05.20	

Студент

Юлія САДОМА

Керівник

Наталія КУЗНЄЦОВА

РЕФЕРАТ

Дипломна робота: 94 с., 76 рис., 14 табл., 2 дод., 7 джерел.

КРЕДИТНИЙ РИЗИК, РИЗК ЛІКВІДНОСТІ, АР, АРКС, АРІКС.

Тема: Аналіз та прогнозування кредитних ризиків у банківській сфері.

В даній роботі досліджується задача аналізу банківських ризиків та їх прогнозування за допомогою регресійних моделей, а саме авторегресії, авторегресії з ковзним середнім, інтегрованої авторегресії з ковзним середнім та множинної лінійної регресії.

Об'єкт дослідження – фінансові показники АТ КБ «ПРИВАТБАНК» з 2001-го року по 2019-го рік, а саме дані про процент кредитного ризику та процент поточної ліквідності.

Предмет дослідження – методи прогнозування часових рядів, а також критерії адекватності математичних моделей і якості прогнозів.

Метою дипломної роботи є створення та реалізація програмного продукту для аналізу та прогнозування кредитних та інших фінансових ризиків, що впливають на стабільність банку.

Результатом роботи є програмний продукт написаний мовою програмування Python. Створений програмний продукт може спростити деякі аспекти роботи та бути корисним у банківській сфері.

ABSTRACT

Diploma work: 94 p., 76 fig., 14 tables, 2 appendixes, 7 sources.

CREDIT RISK, LIQUIDITY RISK, AR, ARMA, ARIMA.

Topic: Analysis and forecasting of credit risks in the banking sector.

The work deals with the problem of analysis of banking risks and their forecasting using regression models, namely autoregression, autoregression with moving average, integrated autoregression with moving average and multiple linear regression.

The object of the study is the financial indicators of JSC CB "PRIVATBANK" from 2001 to 2019, namely the data on the percentage of credit risk and the percentage of current liquidity.

The subject of research - methods of forecasting time series, as well as criteria for the adequacy of mathematical models and the quality of forecasts.

The purpose of the work is to create and implement a software product for analysis and forecasting of credit and other financial risks that affect the stability of the bank.

The result is a software product written in the Python programming language. The created software product can simplify some aspects of work and be useful in the banking sphere.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	9
ВСТУП	10
РОЗДІЛ 1 КРЕДИТНІ РИЗИКИ ТА ПІДХОДИ ДО ЇХ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ	12
1.1 Поняття та класифікація кредитних ризиків.....	12
1.2 Методи оцінки кредитних ризиків	15
1.2.1 Статистичний метод	15
1.2.2 Коефіцієнтний метод	16
1.2.3 VaR метод	17
1.3 Підходи математичного моделювання, що використовуються для прогнозування ризиків.....	18
1.3.1 Регресійний підхід	18
1.3.2 Логістична регресія.....	19
1.3.3 Нечіткі нейронні мережі.....	19
1.4 Огляд існуючих комп'ютерних систем для побудови прогнозу.....	20
1.4.1 Eviews	20
1.4.2 R	22
1.4.3 SAS	23
1.4.4 MatLab	24
1.4.5 Порівняння комп'ютерних систем для побудови моделей фінансово-економічних процесів	26
1.5 Висновки до розділу та постановка задачі дослідження	27
РОЗДІЛ 2 ВИБІР І ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ ТА АНАЛІЗУ ДАНИХ	28
2.1 Основні поняття	28
2.2 Методи заповнення пропусків даних.....	31

2.2.1.Метод випадкового лісу	31
2.2.2 Метод k-найближчих сусідів	31
2.3 Метод фільтрації даних	32
2.4 Моделі для прогнозування	33
2.4.1Авторегресія	33
2.4.2Авторегресія з ковзним середнім	34
2.4.3Модель множинної регресії	35
2.4.4.Сезонна модель ARIMA (SARIMA).....	36
2.5 Критерії адекватності математичних моделей і якості прогнозів	37
2.5.1 Критерії адекватності математичних моделей.....	37
2.5.2 Критерії якості прогнозів	38
2.6 Висновки до розділу 2	40
РОЗДІЛ 3 РОЗРОБКА І ЗАСТОСУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДО ПОБУДОВИ МОДЕЛЕЙ ВИБРАНИХ ПРОЦЕСІВ	41
3.1 Аналіз вхідних даних	41
3.2 Виконання обчислювальних експериментів	45
3.3 Висновки до розділу 3	60
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО – ВАРТІСНИЙ АНАЛІЗ ПП	62
4.1 Постановка задачі проектування	62
4.2. Обґрунтування функцій та параметрів програмного продукту	62
4.3 Економічний аналіз варіантів розробки	69
4.4 Висновки до розділу 4	73
ВИСНОВКИ.....	74
ЛІТЕРАТУРА	76
ДОДАТОК А ІЛЮСТРАТИВНИЙ МАТЕРІАЛ.....	77
ДОДАТОК Б ЛІСТИНГ ПРОГРАМИ	86

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

СП – семантичне правило.

ФН – функція належності.

НМ – нейронні мережі.

ПЗ – програмне забезпечення.

ПП – програмний продукт.

ЛР – лінійна регресія.

ОС – операційна система.

АЧР – аналіз часових рядів.

SAS – Statistical Analysis System.

Eviews – Econometric Views.

VaR – Value-at-Risk.

ВВ – випадкова величина.

MSE - Mean square error.

MAE – Mean absolute error.

MAPE – Mean absolute percentage error.

ВСТУП

Основоположним критерієм будь-якого банку – є його кредитна діяльність. Саме це й основна його відмінність від інших небанківських установ. Кредитування – є найприбутковішою операцією у банківському бізнесі. За їх рахунок формується основна частка чистого доходу (прибутку), яка потім буде йти на відрахування до резервних фондів та грошові виплати акціонерам банку. Проте, разом з тим, невиплата кредитів може викликати банкрутство банку. А це, у свою чергу може спровокувати цілу низку банкрутств підприємств, банків і приватних осіб що мають відношення до банку. Тому, останнім часом приділяється все більше уваги аналізу та оцінці кредитних ризиків з боку банків та інших кредитних установ.

Розглядаючи сучасну банківську систему України, можемо помітити, що деякі банки, використовують надзвичайно ризиковану кредитну політику. Дуже часто це може привести до погіршення фінансового стану банку, а з огляду на вищесказане можемо зробити висновок, що від таких дій може постраждати ціла низка інших підприємств, які мають до нього відношення.

Крім того, відомо, що кредитний ризик належить до відкритих ризиків, що у свою чергу значить – він може регулюватись. І хоч повністю позбутися від ризику неможливо, його вчасна оцінка, аналіз та прогнозування може посприяти його мінімізації.

Метою даної дипломної роботи є створення та реалізація програмного продукту для аналізу та прогнозування кредитних та інших фінансових ризиків, що впливають на стабільність банку. Створений ПЗ може спростити деякі аспекти роботи та бути корисним у банківській сфері.

З огляду на вищесказане, можемо зробити висновок, що кредитні ризики є найголовнішою небезпекою будь-якого банку, а їх вчасна оцінка, аналіз, прогнозування та управління ними – важливий аспект стратегії розвитку та

стабільної роботи будь-яких кредитних установ, саме це зумовлює актуальність обраної теми дипломної роботи.

РОЗДІЛ 1 КРЕДИТНІ РИЗИКИ ТА ПІДХОДИ ДО ЇХ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ

1.1 Поняття та класифікація кредитних ризиків

Два найвагоміших аспекти у фінансово-кредитній та виробничій сферах це ризик та прибуток. Ризиком називають деяку невизначену подію чи умову, яка може внести як позитивний так і негативний вплив у роботу та розвиток певної компанії або підприємства, та може призвести до грошових придбань чи втрат. Розрізняють три основних елемента ризику: об'єкт, суб'єкт та джерело [1].

Об'єктом ризику виступає певна керована система, з невідомою ефективністю, умовами роботи і функціонування, стосовно якої приймаються рішення. Суб'єкт ризику виступає особа (це може бути індивід або колектив), що може приймати рішення відносно об'єкта ризику, а також має зацікавленість у результатах управління цим об'єктом. Джерелом ризику є чинники (явища, процеси), через які зумовлюється невизначеність результатів.

Загалом ризики можна розділити на два види: чисті ризики та спекулятивні ризики. Чисті ризики припускають можливість певних втрат чи взагалі відсутність жодного результату. Спекулятивні ризики, окрім можливості втрат, ще допускають отримання позитивного результату.

Фінансові ризики є різновидом спекулятивних. У широкому значенні. фінансові ризики можна визначити, як ризики, що можуть внести деякі придбання чи втрати грошових коштів у будь-яку підприємницьку діяльність. Це ж саме поняття можемо трактувати й у більш вузькому значенні, а саме як одну з частин комерційних ризиків, що характеризується можливістю фінансових втрат через проведення операцій у фінансово-кредитній і біржовій сферах [2].

Фінансові ризики можна поділити на:

- кредитний ризик;
- валютний ризик;
- ризик ліквідності;
- ринковий ризик;
- процентний ризик;
- ризик інфляції;
- ризик неплатоспроможності;
- ризик зниження фінансової стійкості.

Однією з особливостей даного ризику є можливість понести збитки у результаті проведення різного виду дій та операцій у біржовій та фінансово-кредитній сферах, або ж проведення операцій та дій із цінними паперами. Можемо спостерігати, що з поданих значень, в цілому фінансові ризики можемо поділити на наступні категорії: фінансові і кредитні. Проте, єдиної загальної думки щодо такого поділу серед вчених і спеціалістів даної області наразі немає. Найчастіше можна зустріти як у іноземній так і у вітчизняній літературі, що поняття „фінансові ризики” включає в себе і кредитні ризики, або вживається термін „фінансово-кредитні ризики”. До того ж українська та європейська система класифікації страхування не дають чіткого розмежування цих понять [3].

Одним з найголовніших ризиків характерних для банківської діяльності є кредитний ризик.

Кредитний ризик або ж ризик дефолту – це такий тип ризику, з яким зустрічаються усі банки та інші кредитні установи. Це ризик, що з’являється при своєчасному невиконанні учасником-контрагентом (боржником) фінансових зобов’язань, тобто позичальник не відшкодовує свої боргові виплати, що у свою чергу призводить до фінансових втрат банку. Структура кредитного ризику наведена на рисунку 1.1. Комерційні банки, інвестиційні банки, компанії керування активами, приватні фонди та страхові компанії повинні пильно аналізувати притаманні їм кредитні, щоб ефективно та вигідно діяти на ринку.

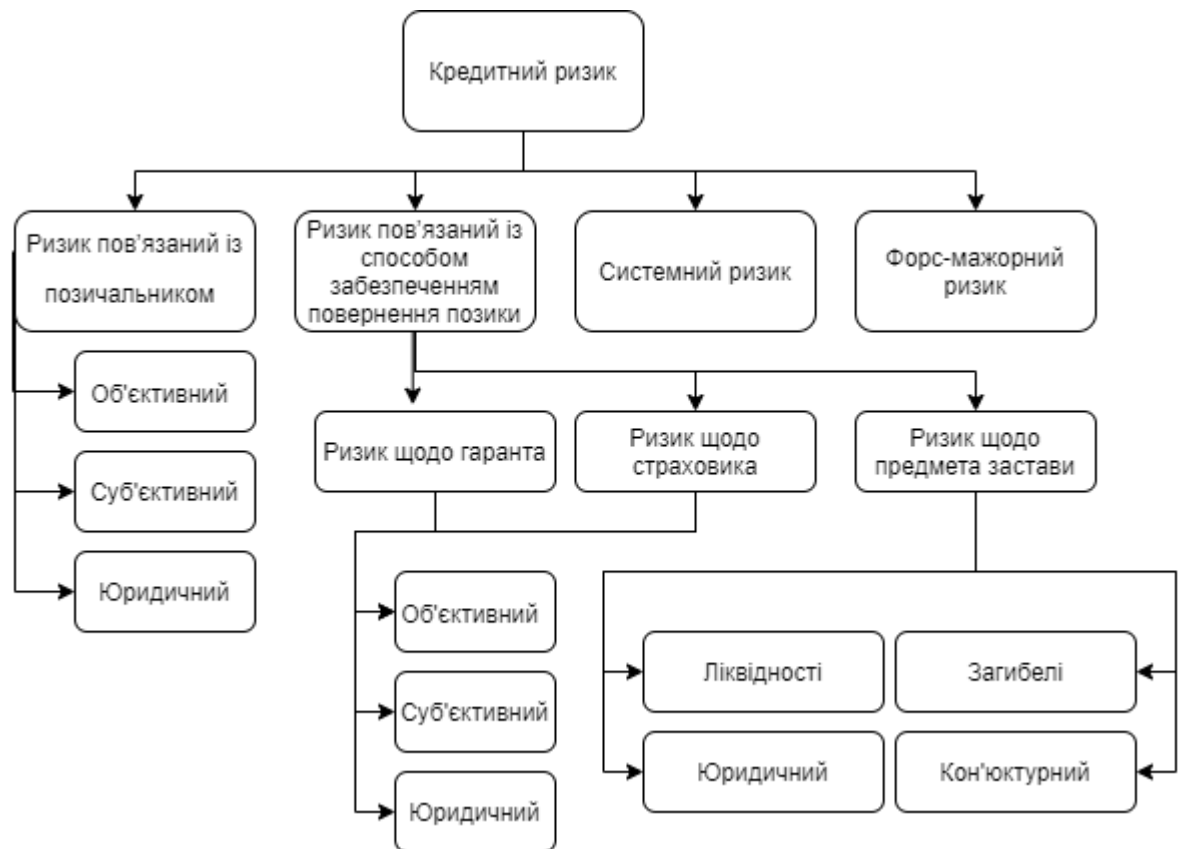


Рисунок 1.1 – Структура кредитного ризику

Вчасне виявлення причин, що свідчать про зростання показника кредитного ризику, та їх аналіз, допомагає банкам та іншим кредитним установам своєчасно приймати рішення щодо подальших дій задля мінімізації цього показника. Комерційні банки, інвестиційні банки, компанії з управління активами, фонди приватного капіталу, фонди венчурного капіталу та страхові компанії повинні проаналізувати кредитні ризики, до яких вони схильні, щоб вигідно діяти на ринку.

Кредитний ризик може з'являтися з чинників, що відносяться як до сторони банку так і до сторони позичальника.

Чинники зі сторони позичальника:

- кредитоспроможність;
- характер укладання кредитної угоди.
- Чинники зі сторони банківських установ:
- зміна кредитної політики;

- концентрація кредитної діяльності у певному сегменті економіки
- надмірна довіра до забезпечення кредитів;
- некоректна організація кредитного процесу.

1.2 Методи оцінки кредитного ризику

1.2.1 Статистичний метод

Якщо банк або кредитна установа володіють достатньою кількістю аналітико-статистичної інформації, доцільно використовувати статистичний метод для оцінки кредитного ризику.

У даному методі оцінки кредитного ризику визначають 4 способи, вони наведені у таблиці 1.1.

Таблиця 1.1 – Статистичний метод оцінки кредитного ризику

Спосіб №	Формула розрахунку	Пояснення значень
1	$R = P(x)$	R – оцінка ризику x – випадкова величина збитку P(x) – вірогідність збитку
2	$R = M(x)$	R – оцінка ризику x – випадкова величина збитку M(x) – математичне сподівання збитку

Продовження таблиці 1.1

Спосіб №	Формула розрахунку	Пояснення значень
3	$R = P(x) \times M(x)$	<p>R – оцінка ризику</p> <p>x – випадкова величина збитку</p> <p>P(x) – вірогідність збитку</p> <p>M(x) – математичне сподівання збитку</p>
4	$R = M(x)$ $\sigma(x) = \sqrt{\sigma^2(x)}$ $K = \frac{\sigma(x)}{R}$	<p>R – оцінка ризику</p> <p>x – випадкова величина збитку</p> <p>M(x) – математичне сподівання збитку</p> <p>$\sigma(x)$ – стандартне (середньоквадратичне) відхилення</p> <p>K – коефіцієнт варіації.</p>

1.2.2 Коефіцієнтний метод

Сенс коефіцієнтного методу полягає у розрахунку відносних характеристик якості кредитного портфеля, а потім у порівнянні їх з нормативними значеннями [4]. У даному методі виконується розрахунок семи коефіцієнтів, за допомогою яких розраховується загальна оцінка кредитного ризику (R), вони подані у таблиці 1.2.

Таблиця 1.2 – Нормативні значення коефіцієнтів

Позначення	Назва	Нормативні значення
K1	Коефіцієнт забезпеченості кредитів	$< 0,9$
K2	Коефіцієнт прострочених позик	0
K3	Коефіцієнт диверсифікації кредитних вкладень	—
K4	Коефіцієнт концентрації кредитів за виданими сумами	$< 0,2$
K5	Коефіцієнт концентрації кредитів за місцем розташування банку	$< 0,1$
K6	Коефіцієнт співвідношення збитку до прибутку по кредитному портфелю	$< 0,7$
K7	Коефіцієнт покриття збитків за позиками	$< 0,2$
R	Загальна оцінка кредитного ризику	$< 0,3$

1.2.3 VaR метод

VaR – універсальний метод оцінки ризиків, який представляє собою статистичну оцінку максимальних втрат внаслідок певної діяльності за деякий проміжок часу.

У випадку нормального розподілу ймовірностей отримуємо дану формулу:

$$\text{VaR} = z_{\alpha} \times \sigma, \quad (1.1)$$

де σ – середньоквадратичне відхилення параметру;

z_{α} – квантиль нормального розподілу ймовірностей.

1.3 Підходи математичного моделювання, що використовуються для прогнозування ризиків

1.3.1 Регресійний підхід

Регресійний підхід – наразі один з найосновніших статистичних методів математичного моделювання певних об’єктів по даних вибірок.

Регресійні моделі розглядають взаємозв’язок між результуючими та пояснюючими змінними. Регресійний аналіз визначає степінь впливу факторів на результуючу ознаку а також показує кількісну оцінку цього впливу. Розробляється рівняння регресії та визначається вплив усіх незалежних змінних на залежну змінну.

В залежності від характеру зв’язку між результуючими та пояснюючими змінним, у регресійному аналізі можуть використовуватись лінійні або ж нелінійні функції. Для визначення цього характеру залежності, зазвичай використовується графічний або табличний метод.[4].

1.3.2 Логістична регресія

Один з різновидів множинної регресії – логістична регресія. У даній моделі розглядається та аналізується зв'язок між декількома незалежними змінними (регресорами) та залежною змінною.

Крім того для прогнозування може застосовуватись й бінарна логістична регресія. Її використання буде доцільним у тих випадках, якщо залежна змінна приймає лише два значення, тобто є бінарною. Таким підходом можемо оцінити ймовірність настання чи ненастання деякої події. Наприклад стабільність/нестабільність банку, повернення кредиту/дефолт тощо.

У цьому випадку застосовується рівняння(функція логіт-перетворення):

$$P = \frac{1}{1+e^{-y}}, \quad (1.2)$$

де P – вірогідність того, що відбудеться деяка подія, яка нас цікавить;

e – основа натуральних логарифмів;

y – деяка змінна, яка визначається рівнянням ЛР (лінійної регресії).

1.3.3 Нечіткі нейронні мережі

Щоб врахувати не лише кількісні а й якісні характеристики, відбувається представлення вектору даних за допомогою лінгвістичних змінних, які визначаються наступним чином:

$$\langle b, T, X, G, M \rangle, \quad (1.3)$$

де b – ім'я даної лінгвістичної змінної;

T – терм-множина змінної, складається з елементів (термів). Загалом терм-множина - це нечітка множина на універсальній множині X ;

G - синтаксичне правило;

M – семантичне правило. СП задає функцію належності нечітких термів. Спершу будуються ФН усіх лінгвістичних змінних, визначаються вид функції та параметри, а потім будується правило G .

У даному підході, на основі нечітких НМ, ФН будується автоматично, а як наслідок автоматично будується і база знань і логічний висновок.

Загальний вигляд алгоритму:

1. Визначається множина вхідних змінних: X .
2. Визначається множина вихідних змінних: Y .
3. Далі формується терм-множина T , яка містить ФН кожної змінної.
4. Формується кінцева узгоджена множина нечітких правил.
5. Визначається істинність для передумов кожного правила.
6. Визначаються нечітких підмножини для вихідних змінних.
7. Виконується композиція нечітких підмножин вихідних змінних.
8. Визначається чітке значення для кожної з лінгвістичних змінних виходу.

1.4 Огляд існуючих комп'ютерних систем для побудови моделей фінансово-економічних процесів

1.4.1 Eviews

Eviews (Econometric Views) – представляє собою пакет потужних інструментів статистики, прогнозування та моделювання за допомогою інноваційного, простого у користуванні та інтуїтивно-зрозумілого об'єктно-орієнтованого інтерфейсу. Програмне забезпечення Eviews розроблене

Quantitative Micro Software (QMS) для операційної системи Windows. ПЗ може використовуватись як для загального статистичного аналізу так і економетричного аналізу даних, а також для оцінки та прогнозування часових рядів. Наразі найновіша версія Eviews - 11, була випущена другого квітня 2019-го року. Інтерфейс Econometric Views показано на рисунках 1.2-1.3.



Рисунок 1.2 – Інтерфейс 1 Econometric Views

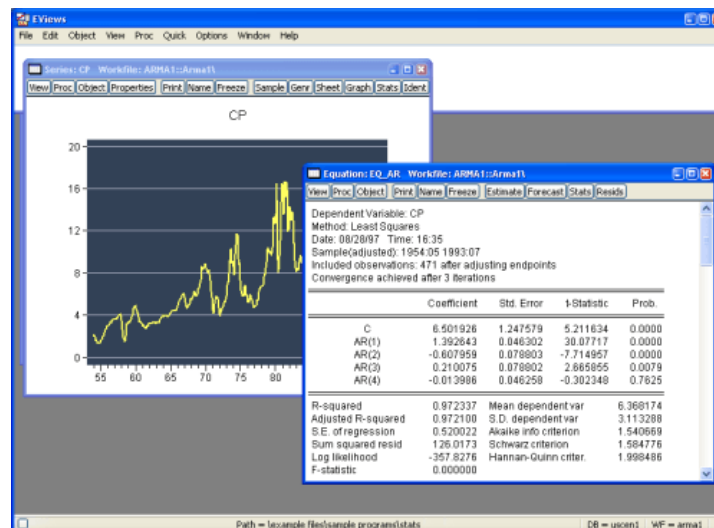


Рисунок 1.3 – Інтерфейс 2 Econometric Views

Переваги:

- достатньо простий у використанні;
- має зручний та зрозумілий інтерфейс.

Недоліки:

- ПЗ Eviews орієнтоване лише на ОС Windows;
- Патентований ПЗ, ціна на нього варіюється від \$1,650 до \$1,995 (В залежності від версії). Проте, у зв'язку з COVID-19 користувачам Eviews було запропоновано безкоштовні тимчасові ліцензії, термін дії яких закінчиться 1 липня 2020 року.

1.4.2 R

R представляє собою мову програмування та обчислювальне середовище для роботи з графікою а також для статистичної обробки даних. Крім того в R реалізовано майже усі актуальні засоби статистичних обчислень, наприклад, такі як АЧР та регресійний аналіз, а також безліч специфічних алгоритмів для вирішення та аналізу аспектів в окремих вузьконаправлених областях. Мова і середовище підтримуються і розвиваються організацією R Foundation. Наразі найновіша версія R — 4.0.0, була випущена 24 квітня 2020-го року. Інтерфейс R показано на рисунку 1.4.

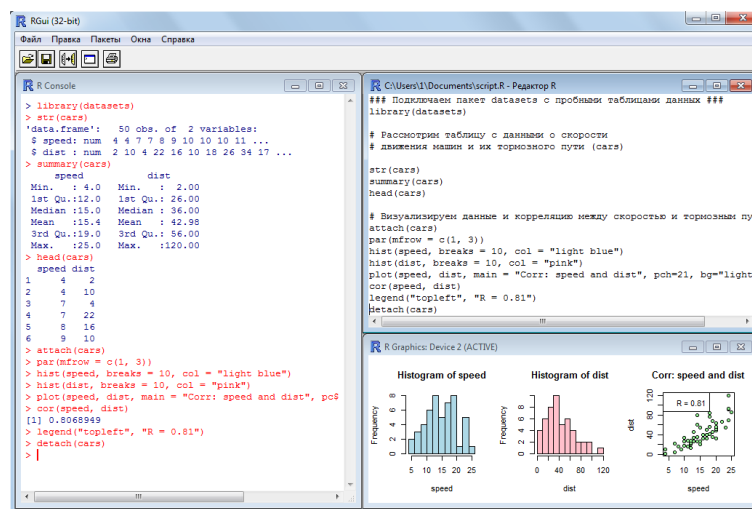


Рисунок 1.4 – Інтерфейс R Studio

Переваги мови програмування R:

- можна встановити на будь-яку операційну систему;
- достатньо простий у вивченні;
- існує безліч додаткових пакетів (більш ніж 7000) для розширення функціоналу;
- повноцінна мова програмування;
- безкоштовне ПЗ у вільному доступі.

Недоліки мови програмування R:

- відносно складний для використання користувачем, який незнайомий з мовами програмування.

1.4.3 SAS

SAS (Statistical Analysis System)- статистичне ПЗ спрямоване на аналіз, обробку, прогнозування та управління масивів даних. ПЗ розроблене Інститутом SAS. Наразі найновіша версія SAS - 9.4, була випущена 10 липня 2013-го року. Інтерфейс SAS продемонстровано на рисунку 1.5.

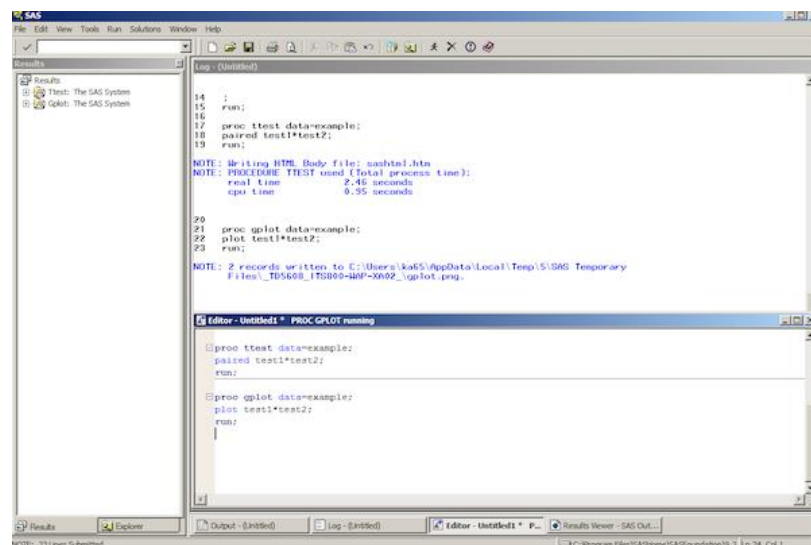


Рисунок 1.5 – Інтерфейс ПЗ SAS

Переваги ПЗ SAS:

- можна встановити на будь-яку операційну систему;
- доволі простий синтаксис, швидкий у вивченні;
- зручний формат вихідних даних;
- інтеграція з базами даних;
- налагодження коду проходить значно простіше, ніж на мові програмування R;

Недоліки ПЗ SAS:

- необхідна покупка, а потім щорічне продовження ліцензії (\$8000.00 на рік);
- SAS значно поступається в продуктивності R та з точки зору об'єму коду;
- документація багатьох алгоритмів SAS не є публічною, тож вивчення мови та роботи ПЗ є сильно обмеженим;

1.4.4 MATLAB

MATLAB (matrix laboratory) – представляє собою чисельне обчислювальне середовище та мову програмування для забезпечення взаємодії користувача з усіма доступними можливостями представлення, оцінки, аналізу та збору даних. Пакет прикладних програм MATLAB було створено компанією The MathWorks. Наразі найновіша версія MATLAB - R2020a, була випущена 19 березня 2020-го року. Інтерфейс MATLAB показано на рисунку 1.6.

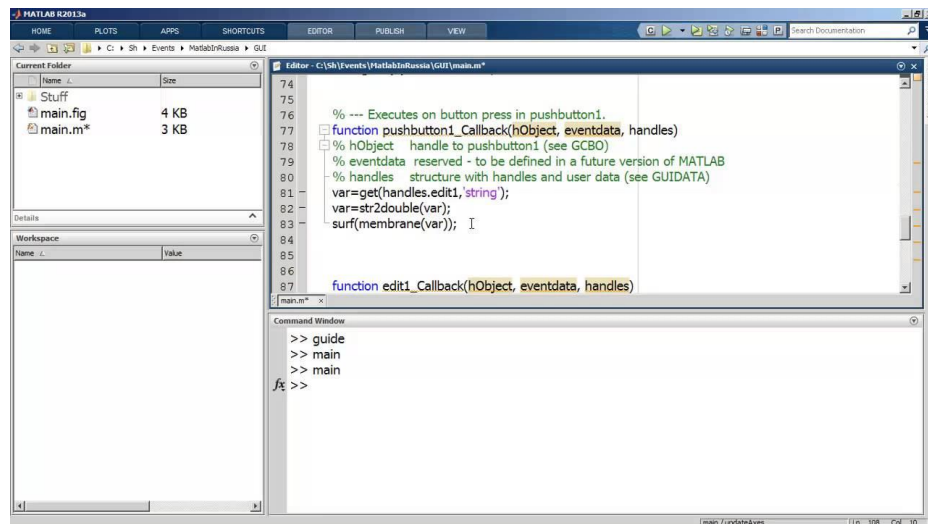


Рисунок 1.6 – Інтерфейс ПЗ MATLAB

Переваги MATLAB:

- доволі простий синтаксис, що дозволяє швидко вивчити мову;
- є можливість перетворювати код MATLAB в код на C або C ++;
- великий спектр різноманітних можливостей.

Недоліки MATLAB:

- необхідна покупка ліцензії на використання, ціна варіюється від \$940 до \$2,350 (В залежності від типу ліцензії);
- вузьконаправлена мова програмування;
- дещо перегружена мова програмування за рахунок операторів, функцій та команд.

1.4.5 Порівняння комп'ютерних систем для побудови моделей фінансово-економічних процесів

У таблиці 1.3 наведено результати порівняльного аналізу вищерозглянутих комп'ютерних систем:

Таблиця 1.3 – Результати порівняльного аналізу ПЗ

Назва ПЗ	Широконаправленість ПЗ	Складність синтаксису та інтерфейсу	ОС	Вартість
Eviews	-	-	Microsoft Windows	\$1,650- \$1,995
R	+	+-	GNU/Linux, Microsoft Windows, macOS BSD	—
SAS	+	-	Microsoft Windows IBM mainframe Unix/Linux OpenVMS Alpha	\$8000.00 на рік
MATLAB	-	-	Microsoft Windows Unix/Linux	\$940- \$2,350

1.5 Висновки до розділу та постановка задачі дослідження

У даному розділі було розглянуто сутність поняття «кредитний ризик», його класифікація. Також було наведено три способи оцінки кредитного ризику та декілька підходів до його прогнозування. Крім того була складена порівняльна таблиця існуючих комп'ютерних систем для побудови моделей фінансово-економічних процесів, з якої можна зробити висновок, що на сьогодні найзручнішим ПЗ а також мовою програмування – є мова R. Метою даної дипломної роботи є створення програмного забезпечення для аналізу та прогнозування кредитних ризиків, тестувати яке будемо на прикладі фінансових даних АТ КБ «ПРИВАТБАНК».

При розробленні відповідного забезпечення потрібно розв'язати наступні завдання:

- огляд та аналіз існуючих методів, що використовуються для прогнозування даних, фільтрації даних та заповнення пропусків даних;
- вибір методу для заповнення пропусків даних;
- вибір методу фільтрації даних;
- вибір та адаптація існуючих методів для прогнозування ризиків;
- розробка програмного продукту на базі обраних математичних методів;
- тестування розробленого програмного забезпечення, порівняння та аналіз результатів.

Вимоги до реалізованого ПП:

- ПП має бути просте у користуванні;
- ПП має коректно працювати на різних вибірках.

РОЗДІЛ 2 МАТЕМАТИЧНІ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ТА АНАЛІЗУ ДАНИХ

2.1 Основні поняття

Розрізняють два різновиди даних для побудови економетричних моделей:

- ті, що описують сукупність процесів в конкретний момент часу t ;
- ті, що описують один певний процес протягом деяких послідовних періодів часу.

По даних першого типу будуються просторові моделі, а по даних другого – моделі часових рядів.

Часовий ряд або ж ряд динаміки $y_1, \dots, y_k, \dots, y_t$ представляє собою значення деякого процесу, об'єкту або ознаки, що вимірюються впродовж певних постійних інтервалів часу t .

Основними компонентами ЧР є

- тренд;
- сезонність;
- цикл;
- похибка;

Розрізняють стаціонарні (рисунок 2.1) та нестаціонарні (рисунок 2.2) ЧР. Стаціонарним називають ряд y_1, \dots, y_T , якщо для будь-якого $\forall s$ розподіл y_t, \dots, y_{t+s} не є залежним від часу t . Інакше кажучи, властивості даного ряду не є залежними від часу.

Нестационарні ЧР не володіють цією властивістю. Варто відмітити, що наявність тренду або наявність сезонності одразу свідчать про нестаціонарність часового ряду, проте з наявності циклів висновки щодо стаціонарності зробити не можемо.

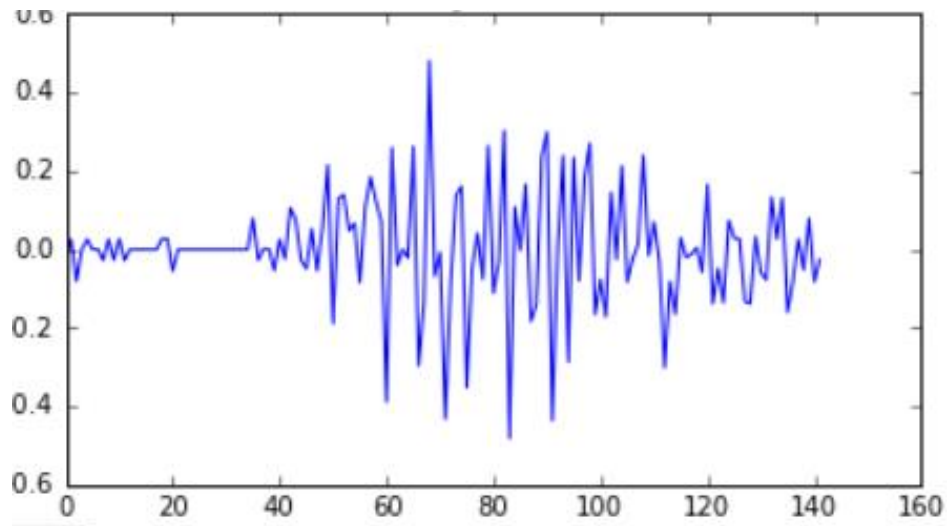


Рисунок 2.1 – Приклад стаціонарного часового ряду

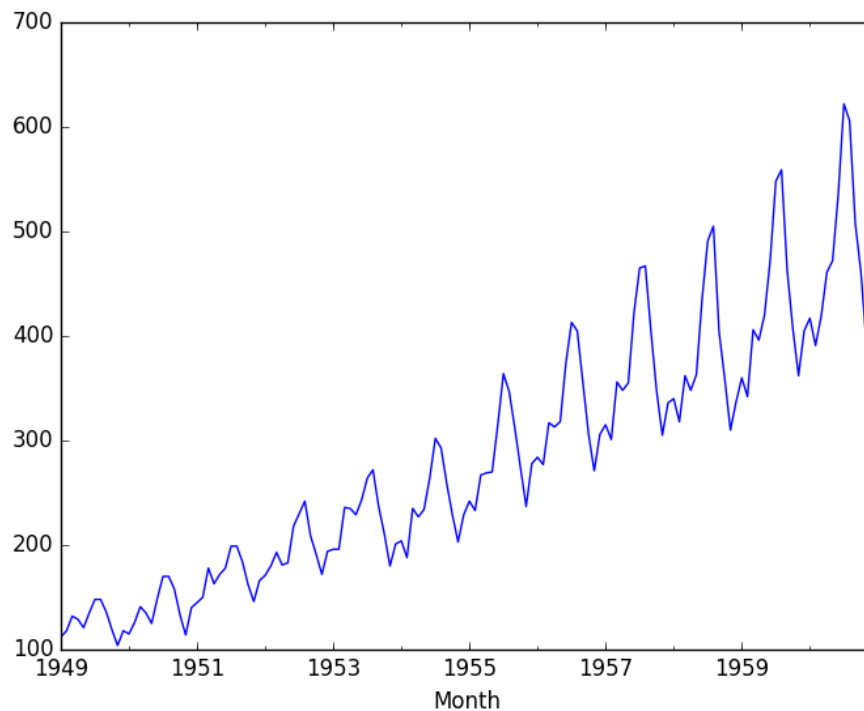


Рисунок 2.2 – Приклад нестационарного часового ряду

На сьогоднішній день існує безліч тестів для перевірки стаціонарності часових рядів.

Тест Діккі-Фуллера (Augmented Dickey-Fuller test) – найвідоміший та один із найпростіших тестів на наявність одиничного кореня. Тест проводиться методом перевірки гіпотез. Нульовою гіпотезою H_0 у тесті Діккі-Фуллера є наявність одиничного кореня, що свідчить про нестационарність

ЧР. Н1 – гіпотеза щодо стаціонарності ЧР. Статистикою даного тесту є класична t-статистика, проте її розподілом виступає вінерівський процес, названий розподілом Діккі-Фуллера.

Розрізняють 3 види тестової регресії:

$$\Delta y_t = b \times y_{t-1} + \varepsilon_t, \quad (2.1)$$

$$\Delta y_t = b_0 + b \times y_{t-1} + \varepsilon_t, \quad (2.2)$$

$$\Delta y_t = b_0 + b_1 \times t + b \times y_{t-1} + \varepsilon_t, \quad (2.3)$$

де Δ – оператор першого диференціювання,

y_t – змінна яку моделюють;

b_0 – константа;

ε_t – похибка.

У першому випадку відсутня константа і тренд. У другому є константа, проте знову відсутній тренд. А у третьому випадку присутні як константа, так і лінійний тренд.

Для кожного з трьох видів тестової регресії існують критичні значення DF-тесту, вони наведені у таблиці 2.1.

Таблиця 2.1 – Критичні значення тесту Діккі-Фуллера

Види тестової регресії	I		II		III	
Об'єм вибірки даних	1%	5%	1%	5%	1%	5%
N=25	-2,66	-1,95	-3,75	-3,00	-4,38	-3,60
N=50	-2,62	-1,95	-3,58	-2,93	-4,15	-3,50
N=100	-2,60	-1,95	-3,51	-2,89	-4,04	-3,45
N=500	-2,58	-1,95	-3,44	-2,87	-3,98	-3,42

2.2 Методи заповнення пропусків даних

2.2.1 Метод випадкового лісу

Пропущені значення даних вибірки ініціалізуються шляхом їх заміни середнім значенням за певною ознакою. Алгоритмом для уточнення пропусків виступає випадковий ліс.

Спершу проводиться навчання на об'єктах без пропусків за певною ознакою. Цей процес виконується для кожної ознаки, що має пропуски значень. Далі відбувається заміна значень даної ознаки для об'єктів, що залишилися, за допомогою алгоритму. Дана дія повторюється до збіжності або до встановленого числа ітерацій.

2.2.2 Метод k-найближчих сусідів

Основним принципом методу k-найближчих сусідів є припущення, що близьким за розташуванням об'єктам властиві схожі ознаки. Тобто, якщо існує об'єкт x з пропущеним значенням, то згідно методу знаходяться об'єкти x_1, \dots, x_k , на основі значень яких визначається значення x за даною формулою:

$$x = \frac{\sum_{k=1}^K x_k}{K}, \quad (2.4)$$

де x – об'єкт з пропущеним значенням;

x_k – об'єкт з відомим значенням.

2.3 Метод фільтрації даних

Фільтром називають алгоритм обробки даних, за допомогою якого можемо позбутися шуму та зайвої інформації.

Експоненціальний фільтр – один із найрозповсюдженіших та найпростіших фільтрів, фільтр першого порядку. Схема роботи експоненціального фільтра показана на рисунку 2.3.

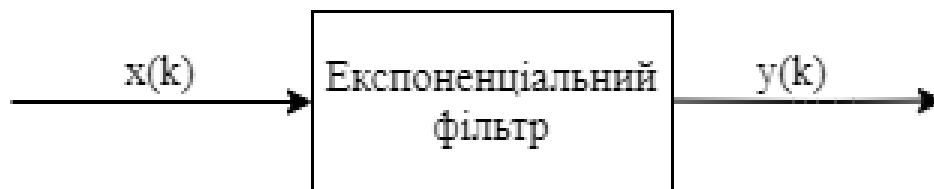


Рисунок 2.3 – Схема роботи експоненціального фільтра

Аналітичне представлення експоненціального фільтра:

$$y_k = a \times x_k + (1 - a) \times y_{k-1} \text{ або } y_k = y_{k-1} + a \times (x_k - y_{k-1}), \quad (2.5)$$

$$a = 1 - e^{\frac{-\Delta T}{\tau}}, \quad (2.6)$$

де x_k – вхідні значення у момент k ;

y_k – відфільтровані вихідні значення у момент k ;

a – "константа згладжування" або коефіцієнт фільтрації;

ΔT – інтервал часу вибірки;

τ – константа часу фільтра.

При збільшенні коефіцієнта a , степінь фільтрації зменшується, і навпаки при зменшенні a – збільшується. Якщо $a = 0$, тоді отримуємо максимальну фільтрацію, при $a = 1$ – мінімальну. На практиці зазвичай використовують значення не більші за 0,5.

Варто відмітити декілька переваг даного фільтра:

- доволі легкий у застосуванні;
- пригнічує різкі скачки вимірів;
- демонструє доволі точні результати прогнозів.

Недоліком експоненціального фільтру є значна затримка.

2.4 Моделі для прогнозування

2.4.1 Авторегресійна модель

Авторегресійна модель позначається як $AR(p)$ та представляє собою стаціонарний процес вигляду:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad (2.7)$$

де ε_t - білий шум, а якщо точніше, то – послідовність незалежних та однаково розподілених ВВ;

a_0 -константа;

p - порядок авторегресії (кількість лагів авторегресії);

q - порядок ковзного середнього (кількість лагів КС);

a_1, a_2, \dots, a_p - авторегресійні параметри моделі.

Проте, для того щоб авторегресійна модель $AR(p)$ залишалась стаціонарною, вводяться певні обмеження щодо її параметрів. Наприклад, для авторегресійної моделі першого порядку $AR(1)$ таким обмеженням є

$$|a_1| < 1, \quad (2.8)$$

де a_1 - авторегресійні параметри моделі.

Загалом, будь-який стаціонарний процес можна подати у вигляді $AR(\infty)$.

2.4.2 Авторегресія з ковзним середнім

Модель авторегресії з ковзним середнім $ARMA(p,q)$ поєднує у собі дві простіші моделі – авторегресії та ковзного середнього, та представляє собою стаціонарний процес вигляду:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}, \quad (2.9)$$

де ε_t – білий шум, а якщо точніше, то – послідовність незалежних та однаково розподілених ВВ;

a_0 – константа;

p – порядок авторегресії (кількість лагів авторегресії);

q – порядок ковзного середнього (кількість лагів КС);

a_1, a_2, \dots, a_p – авторегресійні коефіцієнти моделі;

b_1, b_2, \dots, b_q – авторегресійні коефіцієнти ковзного середнього.

Сума $p+q$ є мінімально можливою. Використовуючи модель $ARMA(p,q)$ ми можемо доволі точно та компактно описати будь-який стаціонарний процес.

Також існує альтернативне подання моделі $ARMA$, використовуючи $L(p,q)$ – оператор затримки:

$$(1 - \sum_{i=1}^p a_i L_i) y_t = (1 - \sum_{j=1}^q b_j L_j) \varepsilon_t, \quad (2.10)$$

$$\text{або } A y_t = B \varepsilon_t, \quad (2.11)$$

де B та A – поліноми виду:

$$B = 1 - \sum_{j=1}^q b_j L_j, \quad (2.12)$$

$$A = 1 - \sum_{i=1}^p a_i L_i, \quad (2.13)$$

де a_1, a_2, \dots, a_p - авторегресійні коефіцієнти моделі;

b_1, b_2, \dots, b_q - авторегресійні коефіцієнти ковзного середнього.

2.4.3 Модель множинної регресії

Процес множинної регресії демонструє кореляційний зв'язок між деякою залежною змінною (y) та декількома незалежними змінними (x_1, x_2, \dots, x_p), тобто

$$y = f(x_1, x_2, \dots, x_p), \quad (2.14)$$

Модель множинної лінійної регресії представляє собою стаціонарний процес вигляду:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_p x_p + \varepsilon, \quad (2.15)$$

де a_0 -константа;

a_1, a_2, \dots, a_p - параметри моделі.

Альтернативний запис:

$$Y = Xa + \varepsilon, \quad (2.16)$$

де X – незалежні змінні;

Y -значення залежних змінних,

a - значення параметрів регресії та вільного члена a_0 ,

ε - вектор значень випадкових залишків.

2.4.4.Сезонна модель ARIMA (SARIMA)

Модель $ARIMA(p,d,q)$ - представляє собою інтегровану модель авторегресії з ковзним середнім $ARMA(p,q)$. Дана модель має вигляд:

$$\Delta^d y_t = a_0 + a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \dots + a_p \Delta y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}, \quad (2.17)$$

де p - порядок авторегресії (кількість лагів авторегресії);

q - порядок ковзного середнього (кількість лагів КС);

a_0 -константа;

a_1, a_2, \dots, a_p - параметри моделі;

b_1, b_2, \dots, b_q - параметри моделі;

ε_t - білий шум;

Δ^d - оператор різниці порядку d .

У випадку, коли $d=0$, з моделі $ARIMA$ отримаємо звичайну модель авторегресії з ковзним середнім $ARMA$.

Модель $SARIMA(p,d,q)(P,D,Q)m$ ж є розширеною версією моделі $ARIMA(p,d,q)$, яка дозволяє не лише обробляти дані з трендом, а й враховувати сезонний компонент ряду.

Перша частина параметрів моделі $SARIMA(p,d,q)(P,D,Q)$ повністю аналогічна моделі $ARIMA(p,d,q)$, тут p - порядок авторегресії, q - порядок ковзного середнього, d – порядок різниці. Проте, у даній моделі ще додаються наступні

параметри: P – сезонний порядок авторегресії, Q – сезонний порядок ковзного середнього, D – порядок сезонної різниці та m – кількість інтервалів одного сезонного періоду.

2.5 Критерії адекватності математичних моделей і якості прогнозів

2.5.1 Критерії адекватності математичних моделей

Математична модель є адекватною, якщо :

- відображає найхарактернішу взаємодію між усіма змінними процесу;
- враховує усі можливі сигнали;
- враховує вплив шумів та зовнішніх збурень;
- враховує обмеження на значення змінних.

AIC (інформаційний критерій Акайке) – критерій, що використовується для вибору найкращої статистичної моделі.

$$AIC = 2k - 2\ln(L), \quad (2.18)$$

де k – кількість параметрів оцінюваної моделі;

L – максимальне значення ф-ції правдоподібності оцінюваної моделі.

R^2 – коефіцієнт детермінації, визначається як рівень залежності варіації залежної змінної від незалежних.

$$R^2 = 1 - \frac{v(y|x)}{v(y)} = 1 - \frac{\sigma^2}{\sigma_y^2}, \quad (2.19)$$

де $V(y | x)$ - умовна дисперсія залежної змінної (дисперсія помилки моделі);

$V(y)$ - дисперсія ВВ y .

Чим більше значення коефіцієнта детермінації тим більша відповідність моделі даним, отже й краща модель. Якщо $R^2 = 1$, тоді між змінними функціональна залежність. На практиці, прийнятними є моделі зі значенням коефіцієнта детермінації більше 0.5.

Критерій Дарбіна-Уотсона використовується для визначення наявності автокореляції залишків першого порядку регресійної моделі.

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \approx 2(1 - p_1), \quad (2.20)$$

де p_1 – коефіцієнт автокореляції першого порядку.

Якщо $DW = 2$ – це говорить про відсутність автокореляції.

У разі, коли DW прямує до 0, можемо казати про наявність позитивної (додатньої) автокореляції, а коли DW прямує до 4 – про наявність негативної (від’ємної) автокореляції.

2.5.2 Критерії якості прогнозів

Якість отриманого прогнозу будемо визначати за наступними критеріями:

MSE (Mean Square Error) – середньоквадратична похибка моделі:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (2.21)$$

де y_i – фактичне (реальне) значення часового ряду;

\hat{y}_i – значення прогнозу.

RMSE (Root Mean Square Error) корінь з середньоквадратичної похибки:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (2.22)$$

де y_i – фактичне (реальне) значення часового ряду;

\hat{y}_i – значення прогнозу.

MAE (Mean absolute error) – середня абсолютна похибка:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (2.23)$$

де y_i – фактичне (реальне) значення часового ряду;

\hat{y}_i – значення прогнозу.

MAPE (Mean absolute percentage error)- середня абсолютна похибка в процентах:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\%, \quad (2.24)$$

де y_i – фактичне (реальне) значення часового ряду;

\hat{y}_i – значення прогнозу.

Theil – коефіцієнт Тейла:

$$\text{Theil} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^T (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^T (y_i)^2 + \frac{1}{N} \sum_{i=1}^T (\hat{y}_i)^2}}, \quad (2.25)$$

де y_i – фактичне (реальне) значення часового ряду;

\hat{y}_i – значення прогнозу.

Якщо коефіцієнт Тейла $Theil = 1$ – це говорить про те, що побудована модель не підходить для прогнозування.

У разі, коли коефіцієнт Тейла дорівнює 0, можемо казати про ідеальність побудованої моделі.

2.6 Висновки до розділу 2

У даному розділі були розглянуті основні поняття та характеристики часового ряду. Наведені статистичні тести, що дозволяють провести перевірку часового ряду на стаціонарність. Також були розглянуті методи до заповнення пропусків – метод випадкового лісу, метод k-найближчих сусідів, та підхід до фільтрації даних у вигляді експоненціального фільтру. Також було наведено чотири математичні моделі для опису та прогнозування ЧР. Крім того було наведено критерії, що дозволяють оцінити адекватність відповідної математичної моделі та якість прогнозу.

РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ТА ПРОГНОЗІВ ДЛЯ ОБРАНИХ ПРОЦЕСІВ

3.1 Аналіз вхідних даних

Практичною частиною моєї бакалаврської роботи була розробка та реалізація програмного забезпечення (продукту) для побудови та аналізу вищезазначених моделей прогнозування часових рядів, а також для аналізу та коригування вхідних даних.

У даному розділі будемо працювати та проводити експерименти з фінансовими даними АТ КБ «ПРИВАТБАНК», що були отримані з фінансових та статистичних звітностей з офіційного сайту банку та з фінансового порталу Finance.ua. Дані представлені у вигляді поквартальних фінансових показників та охоплюють часовий проміжок з 2001-го року по 2019-го рік. Для подальшої роботи було обрано наступні фінансові показники:

- 1) Чистий прибуток (Net income) – підсумковий фінансовий результат роботи банку, після сплати усіх податків та платежів.
- 2) Активи банку (Asset) – фінансові ресурси банку до яких відносяться як власні так і позикові кошти.
- 3) Зобов'язання (Liabilities) банку.
- 4) Розмір кредитного портфеля (Loan portfolio) – розмір усіх непогашених позичальником кредитів на деяку певну дату.
- 5) Депозити фізичних осіб (Retail deposit)- грошові вкладення фізичних осіб.
- 6) Розмір кредитного ризику на одного контрагента (H7) – показник характеризується співвідношенням загальної суми вимог та зобов'язань до регулятивного капіталу відносно контрагента. Згідно з нормативами, встановленими Національним банком України, даний показник кредитного ризику не повинен перевищувати 25%.

- 7) Показник поточної ліквідності (Н5) – характеризується співвідношенням оборотних активів до поточних зобов'язань. Згідно з нормативами, встановленими Національним банком України, даний показник ліквідності повинен бути не менше 40%.

Побудуємо графіки семи показників та наведемо загальну інформацію про ЧР (рисунки 3.1–3.14)



Рисунок 3.1 – Графік зміни розміру чистого прибутку за досліджуваний період

min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-135309.00	79.75	396.00	-235.30	991.50	32609.00

Рисунок 3.2 – Загальна інформація про ЧР (Net income)



Рисунок 3.3 – Графік зміни розміру зобов'язань за досліджуваний період

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5460	20282	89737	112809	214281	259577

Рисунок 3.4 – Загальна інформація про ЧР (Liabilities)

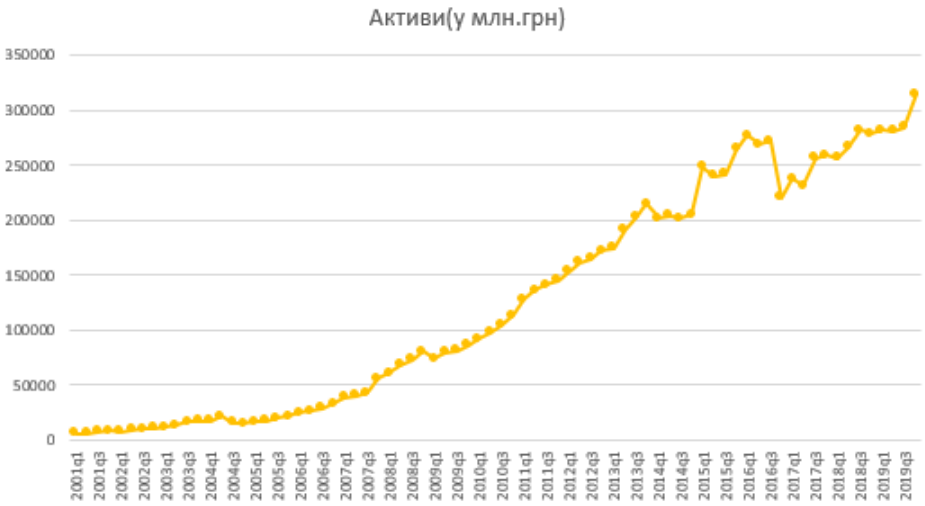


Рисунок 3.5 – Графік зміни розміру активів банку за досліджуваний період

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5742	21673	100933	126706	222752	314106

Рисунок 3.6 – Загальна інформація про ЧР (Asset)

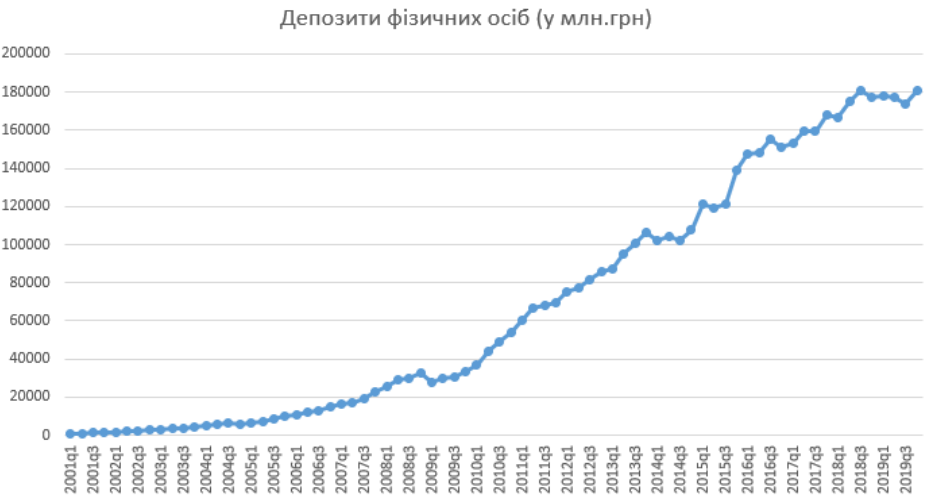


Рисунок 3.7 – Графік зміни розміру депозитів фізичних осіб за досліджуваний період

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
898	9603	46540	68074	119714	181013

Рисунок 3.8 – Загальна інформація про ЧР (Retail deposit)



Рисунок 3.9 – Графік зміни розміру кредитного портфелю банку за досліджуваний період

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1995	16508	55976	67707	106541	186706

Рисунок 3.10 – Загальна інформація про ЧР (Loan portfolio)



Рисунок 3.11 – Графік зміни проценту максимального кредитного ризику на одного контрагента за досліджуваний період

min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.47	17.30	20.52	20.07	22.14	92.17

Рисунок 3.12 – Загальна інформація про ЧР (Н7)



Рисунок 3.3 – Графік зміни проценту поточної ліквідності банку за досліджуваний період

min.	1st Qu.	Median	Mean	3rd Qu.	Max.
45.84	68.39	74.70	85.51	89.62	176.11

Рисунок 3.14 – Загальна інформація про ЧР (Н5)

Загалом бачимо, що суттєві стрибки та зміни показників припадають на період банківської кризи та націоналізації ПриватБанку.

3.2 Аналіз та прогнозування досліджуваних часових рядів

Найпершим кроком виконуємо заповнення пропусків даних у вибірці. Для більш зручної роботи з пропусками, замінимо їх спеціальним значенням. У нашому випадку значення буде дорівнювати «-1» (рисунок 3.15).

	B	C	D	E	F	G	H	I
1	Date	Net income	Assets	Liabilities	Loan_portfolio	Retail dep	H7	H5
2	2001q1	7	5742	5460	1995	898	24.6072	-1
3	2001q2	15	6271	5980	2442	1061	20.5312	-1
4	2001q3	30	7627	7316	2828	1200	18.5771	72.3565
5	2001q4	34	8254	7890	3365	1408	20.5771	70.5247
6	2002q1	8	8036	7664	3366	1716	20.6072	74.7045
7	2002q2	19	9090	8690	3958	1954	20.5312	74.8235
8	2002q3	30	10248	9837	4386	2317	21.5374	67.21
9	2002q4	148	10858	10311	4437	2615	19.6052	70.4715
10	2003q1	7	11551	10996	5279	3123	20.6072	74.9776
11	2003q2	21	13824	13255	5951	3380	17.5154	80.22
12	2003q3	34	16784	16202	7019	3800	20.5771	71.7654
13	2003q4	60	17490	16534	7502	4287	22.6196	70.8145
14	2004q1	14	17874	16904	8611	4967	19.5312	75.3968
15	2004q2	143	21676	20577	11010	5422	-1	-1
16	2004q3	236	15589	14397	11249	6170	-1	-1
17	2004q4	163	14668	13204	11109	5558	-1	-1
18	2005q1	205	16469	-1	12715	6585	20.562	70.0342
19	2005q2	215	17808	15916	13724	7091	23.4939	70.0948

Рисунок 3.15 – Частина вибірки із заміненними спеціальним значенням пропусками

Для заповнення пропуск обираємо метод випадкового лісу, та отримуємо наступні результати (рисунок 3.16).

	B	C	D	E	F	G	H	I
1	Date	Net income	Assets	Liabilities	Loan_portfolio	Retail dep	H7	H5
2	2001q1	7	5742	5460	1995	898	24.6072	74.7045
3	2001q2	15	6271	5980	2442	1061	20.5312	74.8235
4	2001q3	30	7627	7316	2828	1200	18.5771	72.3565
5	2001q4	34	8254	7890	3365	1408	20.5771	70.5247
6	2002q1	8	8036	7664	3366	1716	20.6072	74.7045
7	2002q2	19	9090	8690	3958	1954	20.5312	74.8235
8	2002q3	30	10248	9837	4386	2317	21.5374	67.21
9	2002q4	148	10858	10311	4437	2615	19.6052	70.4715
10	2003q1	7	11551	10996	5279	3123	20.6072	74.9776
11	2003q2	21	13824	13255	5951	3380	17.5154	80.22
12	2003q3	34	16784	16202	7019	3800	20.5771	71.7654
13	2003q4	60	17490	16534	7502	4287	22.6196	70.8145
14	2004q1	14	17874	16904	8611	4967	19.5312	75.3968
15	2004q2	143	21676	20577	11010	5422	20.6052	69.9849
16	2004q3	236	15589	14397	11249	6170	20.4939	70.32
17	2004q4	163	14668	13204	11109	5558	20.562	70.535
18	2005q1	205	16469	14587	12715	6585	20.562	70.0342
19	2005q2	215	17808	15916	13724	7091	23.4939	70.0948

Рисунок 3.15 – Частина вибірки із заповненими пропусками методом випадкового лісу

Далі виконаємо процедуру фільтрації з використанням експоненціального фільтру, щоб позбутися шуму та зайвої інформації (рисунки 3.16-3.17). Коефіцієнт фільтрації θ обираємо рівним 0.3.

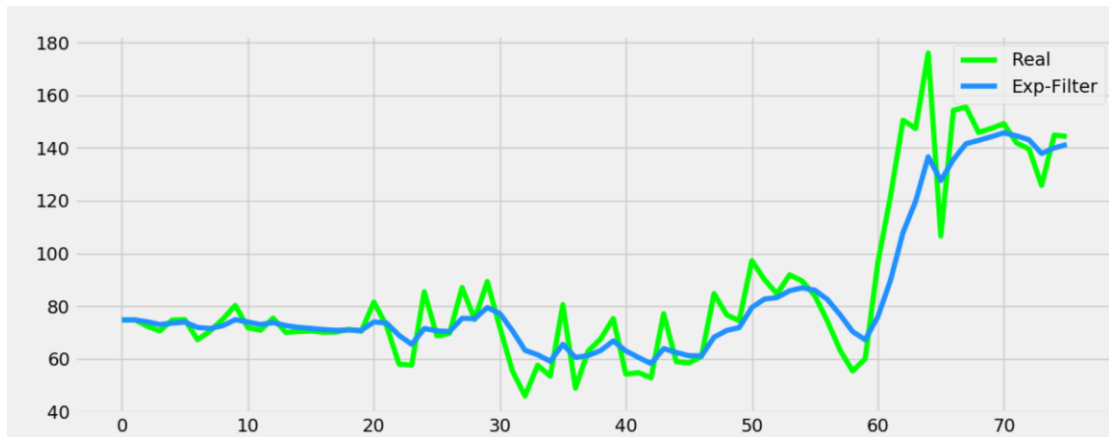


Рисунок 3.16 – Графік даних поточної ліквідності (Н5) з використанням експоненціального фільтру

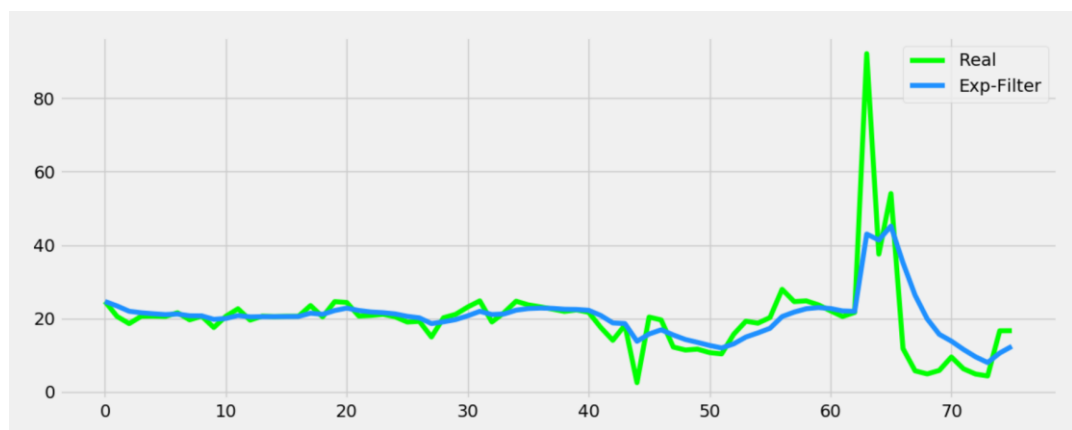


Рисунок 3.17 – Графік даних кредитного ризику (Н7) з використанням експоненціального фільтру.

Надалі будемо працювати з фільтрованими даними

Виконаємо перевірку рядів на стаціонарність.

Спершу проведемо тест Дікі-Фуллера для ряду Н7 (процент максимального кредитного ризику на одного контрагента) (рисунок 3.18).

```
-----
adf: -3.8938083703352064
p-value: 0.0020808895742416905
Critical values: {'1%': -3.5232835753964475, '5%': -2.902030597326081, '10%': -2.5883710883843123}
одиначних коренів немає - ряд стаціонарний
-----
```

Рисунок 3.18 – Результат тесту Дікі-Фуллера для ряду Н7

Значення коефіцієнта p -value досить мале. Це свідчить про те, що ми відхиляємо нульову гіпотезу H_0 про наявність одиничного кореня, та приймаємо альтернативну гіпотезу. Тож маємо стаціонарний ряд.

Побудуємо АКФ та ЧАКФ даного ряду (рисунок 3.19).

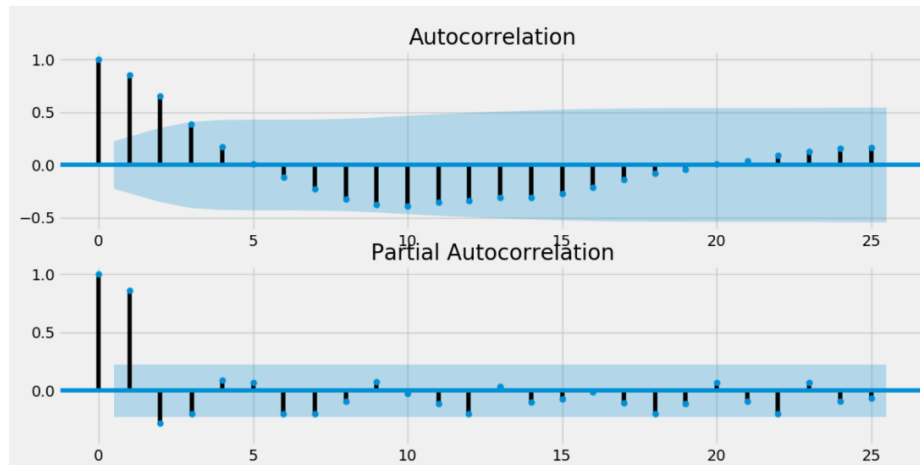


Рисунок 3.19 – АКФ та ЧАКФ ряду Н7

Почнемо побудову моделей з моделей нижчих порядків. Такі моделі доволі часто показують гарну адекватність та прогноз високої якості.

Будуємо модель AR(1) (рисунок 3.20):

ARMA Model Results						
=====						
Dep. Variable:	y	No. Observations:	75			
Model:	ARMA(1, 0)	Log Likelihood	-193.799			
Method:	css-mle	S.D. of innovations	3.177			
Date:	Mon, 01 Jun 2020	AIC	393.599			
Time:	03:37:34	BIC	400.551			
Sample:	0	HQIC	396.375			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	20.1243	2.503	8.038	0.000	15.218	25.031
ar.L1.y	0.8642	0.058	14.996	0.000	0.751	0.977

Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	1.1572	+0.0000j	1.1572	0.0000		

R-square: 0.853560						
Durbin-Watson: 1.711						

Рисунок 3.20 – Результати оцінки моделі AR(1)

Отримуємо наступні характеристики адекватності моделі:

Коефіцієнт детермінації $R^2 = 0.853560$, що є доволі непоганим результатом, критерій Дарбіна-Уотсона DW становить 1.711. Тож, можемо сказати, що отримана модель є досить адекватною. Інформаційний критерій Акайке $AIC=393.59$, він нам знадобиться для порівняння даної моделі з іншими.

Виконавши однокроковий прогноз, отримуємо характеристики його якості (рисунок 3.21).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.07936836	3.156664	1.497191	-2.388377	7.434088	0.9586575

Рисунок 3.21 – Характеристики якості однокрокового прогнозу AR(1)

Будуємо модель AR(2) (рисунок 3.22):

ARMA Model Results						
=====						
Dep. Variable:	y	No. Observations:	75			
Model:	ARMA(2, 0)	Log Likelihood	-190.318			
Method:	css-mle	S.D. of innovations	3.165			
Date:	Mon, 01 Jun 2020	AIC	388.636			
Time:	12:51:26	BIC	397.852			
Sample:	0	HQIC	392.312			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	-0.1894	0.544	-0.348	0.000	-1.255	0.876
ar.L1.y	0.1186	0.113	1.049	0.000	-0.103	0.340
ar.L2.y	0.2119	0.113	1.879	0.060	-0.009	0.433
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	1.9104	+0.0000j	1.9104	0.0000		
AR.2	-2.4699	+0.0000j	2.4699	0.5000		

R-square: 0.752142						
Durbin-Watson: 1.991238						

Рисунок 3.22 – Результати оцінки моделі AR(2)

Маємо наступні характеристики адекватності моделі:

Коефіцієнт детермінації $R^2 = 0.752142$, що є гіршим результатом, порівняно з попередньою моделлю, однак можемо спостерігати покращення критерію Дарбіна-Уотсона $DW = 1.991$. Інформаційний критерій Акайке становить 388.636.

Виконавши однокроковий прогноз, отримуємо характеристики його якості (рисунок 3.23).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.0524857	3.055489	1.388911	-2.174961	6.76876	0.8893256

Рисунок 3.23 – Характеристики якості однокрокового прогнозу AR(2)

Можемо зробити висновок, що у порівнянні з попередньою моделлю, дана модель краще підходить для прогнозування.

Далі будуюмо модель ARMA(1, 2) (рисунок 3.24).

ARMA Model Results						
=====						
Dep. Variable:	y	No. Observations:	75			
Model:	ARMA(1, 2)	Log Likelihood	-186.103			
Method:	css-mle	S.D. of innovations	2.856			
Date:	Mon, 01 Jun 2020	AIC	382.207			
Time:	03:54:39	BIC	393.794			
Sample:	0	HQIC	386.833			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	20.4293	1.937	10.548	0.000	16.633	24.226
ar.L1.y	0.7185	0.094	7.658	0.000	0.535	0.902
ma.L1.y	0.2696	0.124	2.180	0.029	0.027	0.512
ma.L2.y	0.4529	0.099	4.553	0.000	0.258	0.648
Roots						

	Real	Imaginary	Modulus	Frequency		

AR.1	1.3919	+0.0000j	1.3919	0.0000		
MA.1	-0.2977	-1.4559j	1.4860	-0.2821		
MA.2	-0.2977	+1.4559j	1.4860	0.2821		

R-square: 0.98564						
Durbin-Watson: 1.993865						

Рисунок 3.24 – Результати оцінки моделі ARMA(1,2)

У цьому випадку спостерігаємо покращення характеристик адекватності моделі:

Коефіцієнт детермінації збільшився до $R^2 = 0.98564$, критерій Дарбіна-Уотсона становить $DW = 1.993865$ та наближається до свого найкращого значення. Інформаційний критерій Акайке у цьому випадку найменший, порівняно з попередніми моделями, та становить 382.207.

Характеристики якості однокрокового прогнозу наступні (рисунок 3.25).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.05053761	2.83752	1.238694	-1.93763	6.414319	0.7931411

Рисунок 3.25 – Характеристики якості однокрокового прогнозу ARMA(1,2)

Дана модель є найбільш прийнятною для побудови прогнозу, порівняно з попередніми.

Будуємо модель ARMA(2,1) (рисунок 3.26).

ARMA Model Results						
=====						
Dep. Variable:	y	No. Observations:	75			
Model:	ARMA(2, 1)	Log Likelihood	-189.543			
Method:	css-mle	S.D. of innovations	3.131			
Date:	Mon, 01 Jun 2020	AIC	389.086			
Time:	12:58:24	BIC	400.607			
Sample:	0	HQIC	393.682			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	-0.1865	0.528	-0.353	0.000	-1.222	0.849
ar.L1.y	-0.2027	0.251	-0.807	0.000	-0.695	0.290
ar.L2.y	0.2877	0.111	2.592	0.010	0.070	0.505
ma.L1.y	0.3403	0.246	1.385	0.060	-0.141	0.822
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	-1.5450	+0.0000j	1.5450	0.5000		
AR.2	2.2496	+0.0000j	2.2496	0.0000		
MA.1	-2.9387	+0.0000j	2.9387	0.5000		

R-square: 0.847						
Durbin-Watson: 1.694						

Рисунок 3.26 – Результати оцінки моделі ARMA(2,1)

Характеристики адекватності моделі є наступними:

$R^2 = 0.847$; $DW = 1.694$, вони є значно гіршими, ніж у попередній моделі.

Інформаційний критерій Акайке збільшився та становить 389.086.

Характеристики якості однокрокового прогнозу наступні (рисунок 3.27).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.1609428	2.875344	1.508952	-0.5911706	7.218115	0.9661878

Рисунок 3.27 – Характеристики якості однокрокового прогнозу ARMA(2,1)

Можемо зробити висновок, що у порівнянні з попередньою моделлю, дана модель гірше підходить для прогнозування.

Побудуємо модель множинної регресії (рисунок 3.28), та порівняємо її характеристики з попередніми моделями.

OLS Regression Results						
Dep. Variable:	Crexp	R-squared:	0.603			
Model:	OLS	Adj. R-squared:	0.574			
Method:	Least Squares	F-statistic:	21.24			
Date:	Mon, 01 Jun 2020	Prob (F-statistic):	7.32e-13			
Time:	12:44:16	Log-Likelihood:	-211.75			
No. Observations:	76	AIC:	435.5			
Df Residuals:	70	BIC:	449.5			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	15.0070	5.425	2.766	0.007	4.188	25.826
Liabilities	3.301e-05	8.99e-05	0.367	0.715	-0.000	0.000
Deposit	-1.375e-05	0.000	-0.116	0.908	-0.000	0.000
Loan_portfolio	-4.625e-05	3.86e-05	-1.199	0.234	-0.000	3.07e-05
Liquidity	0.4222	0.043	9.871	0.000	0.337	0.508
LiquidityExp	-0.0322	0.071	-0.451	0.653	-0.175	0.110
Omnibus:	29.194	Durbin-Watson:	1.497			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	93.600			
Skew:	1.110	Prob(JB):	4.73e-21			
Kurtosis:	7.962	Cond. No.	2.17e+06			

Рисунок 3.28 –Результати оцінки моделі OLS

У цьому випадку спостерігаємо погіршення характеристик адекватності моделі, порівняно з чотирма попередніми випадками: $R^2 = 0.603$; $DW = 1.497$.

Виконавши однокроковий прогноз, отримуємо наступні характеристики його якості: $MAE = 2.42254$; $MAPE = 13.32067$; $RMSE = 3.924754$.

Можемо зробити висновок, що у порівнянні з попередніми моделями, дана найгірше підходить для прогнозування.

Занесемо всі отримані результати у таблицю 3.1:

Таблиця 3.1 – Характеристики побудованих моделей

Model	MAE	MAPE	RMSE	R^2	DW	AIC
AR(1)	1.497191	7.434088	3.156664	0.853560	1.711	393.59
AR(2)	1.388911	6.76876	3.055489	0.752142	1.991	394.93
ARMA(1,2)	1.238694	6.414319	2.83752	0.98564	1.994	382.20
ARMA(2,1)	1.508952	7.218115	2.875344	0.847	1.694	389.086
OLS	2.42254	13.32067	3.924754	0.603	1.497	435.5

Порівнявши показники усіх моделей, для побудови прогнозу обираємо модель ARMA(1,2). Спрогнозуємо процент кредитного ризику на одного

контрагента на 8 кроків, тобто на чотири квартали 2020-го року та на чотири квартали 2021-го року (рисунок 3.29).

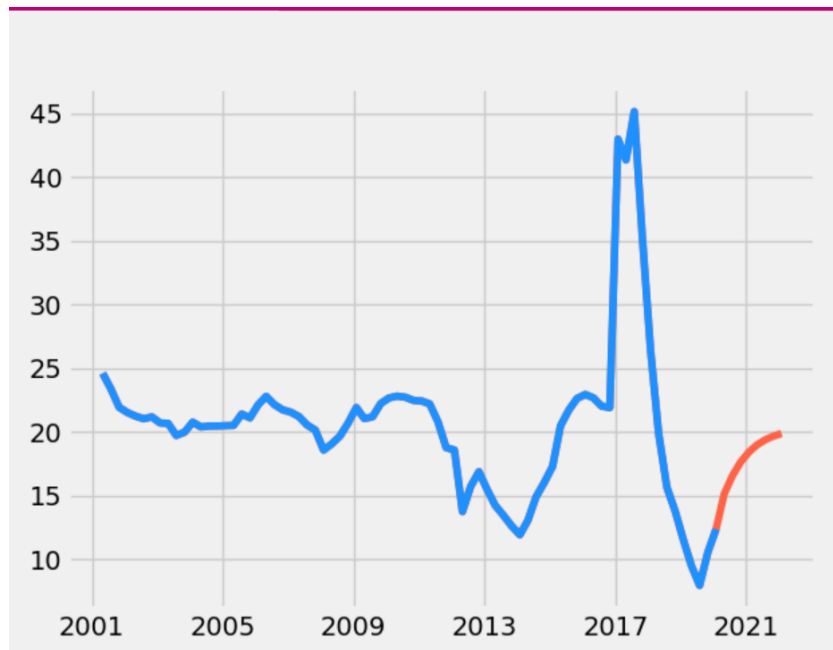


Рисунок 3.29 –Результати прогнозування ряду Н7 на 8 кроків вперед
Результати прогнозування наведемо у таблиці 3.2.

Таблиця 3.2 – Результати прогнозування ряду Н7

Дата	Значення прогнозу
2020q1	15.12625538
2020q2	16.55940161
2020q3	17.64218492
2020q4	18.42119572
2021q1	18.98155416
2021q2	19.38457078
2021q3	19.67462327
2021q4	19.88344807

Загалом, згідно прогнозу, можемо зробити висновок, що процент кредитного ризику на одного контрагента продовжить зростати, проте не перевищує норматив 25%, тож банку нема необхідності переглядати стратегію дій, щодо регулювання кредитного ризику.

Далі будемо досліджувати ряд Н5. Проведемо тест Дікі-Фуллера для ряду Н5 (розмір поточної ліквідності), результати якого наведено на рисунку 3.30.

```
-----
adf: -0.5788915874405699
p-value: 0.8756806856331416
Critical values: {'1%': -3.5232835753964475, '5%': -2.902030597326081, '10%': -2.5883710883843123}
є одиничні корені - ряд не стаціонарний
-----
```

Рисунок 3.30 – Результати тесту Дікі-Фуллера для ряду ряду Н5
Значення коефіцієнта p-value досить високе, що свідчить про те, що ми приймаємо нульову гіпотезу H_0 про наявність одиничного кореня. Наразі маємо нестационарний ряд. Продиференціюємо даний ряд та отримаємо ряд перших різниць (рисунок 3.31).

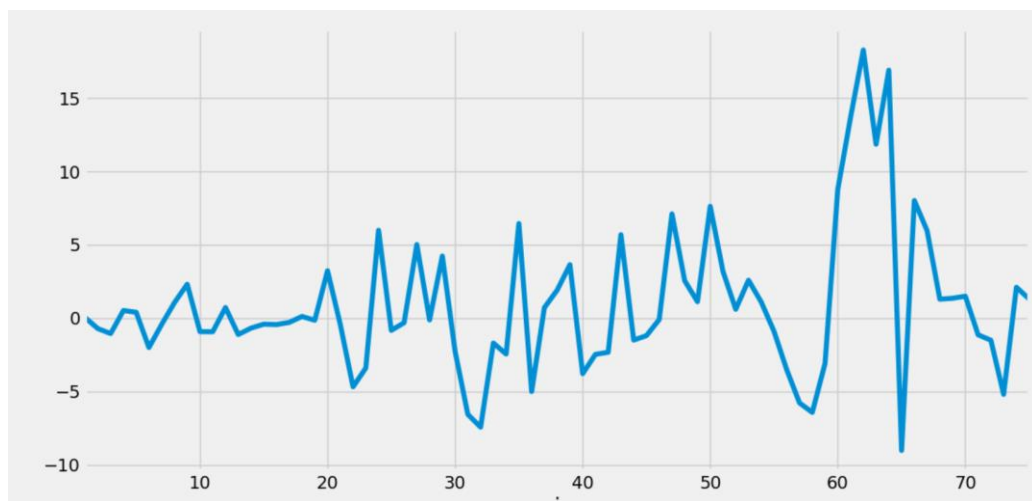


Рисунок 3.31 – Ряд перших різниць ряду поточної ліквідності (Н5)
Проведемо тест Дікі-Фуллера для даного ряду (рисунок 3.32).

```
-----
adf: -12.253428541451338
p-value: 9.426903951848515e-23
Critical values: {'1%': -3.5219803175527606, '5%': -2.9014701097664504, '10%': -2.58807215485756}
одиничних коренів немає - ряд стаціонарний
-----
```

Рисунок 3.32 – Результати тесту Дікі-Фуллера для ряду перших
різниць

Значення коефіцієнта p -value значно мале, що свідчить про те, що ми відхиляємо нульову гіпотезу H_0 про наявність одиничного кореня, та приймаємо альтернативну гіпотезу. Тож одержаний ряд перших різниць є стаціонарним.

Побудуємо АКФ та ЧАКФ для ряду перших різниць (рисунок 3.33).

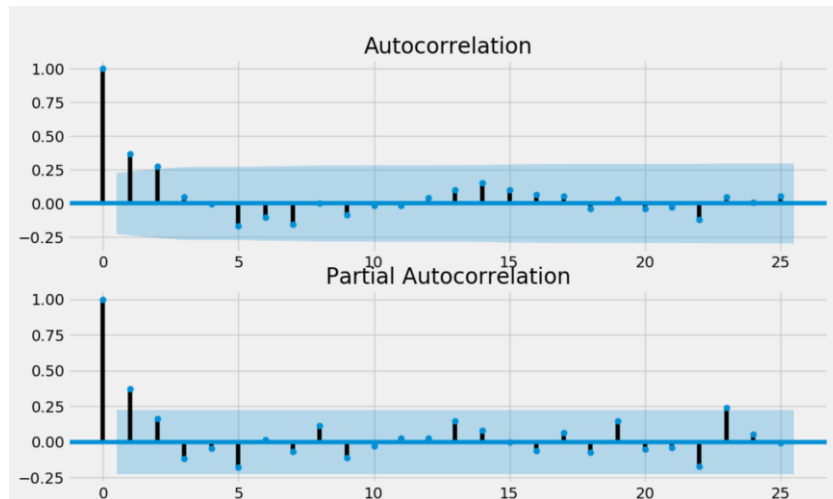


Рисунок 3.33 – АКФ та ЧАКФ ряду $H5$

Будуємо модель $ARIMA(0,1,0)$ (рисунок 3.34):

ARIMA Model Results						
Dep. Variable:	D.y	No. Observations:	74			
Model:	ARIMA(0, 1, 0)	Log Likelihood	-223.482			
Method:	css	S.D. of innovations	4.958			
Date:	Mon, 01 Jun 2020	AIC	450.963			
Time:	04:06:23	BIC	455.571			
Sample:	1	HQIC	452.801			
	coef	std err	z	P> z	[0.025	0.975]
const	0.8829	0.576	1.532	0.126	-0.247	2.013
R-square: 0.9321						
Durbin-Watson: 1.738						

Рисунок 3.34 – Результати оцінки моделі $ARIMA(0, 1, 0)$

Отримуємо наступні характеристики адекватності моделі:

Коефіцієнт детермінації $R^2 = 0.9321$, що є гарним результатом, критерій Дарбіна-Уотсона DW становить 1.738. Тож, можемо сказати, що отримана модель є доволі адекватною. Інформаційний критерій Акайке $AIC=450.963$, він нам знадобиться для порівняння даної моделі з іншими.

Характеристики якості однокрокового прогнозу наступні (рисунок 3.35).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.8778014	4.971936	3.305246	0.6797924	3.94246	0.9871357

Рисунок 3.35 – Характеристики якості однокрокового прогнозу ARIMA(0,1,0)

Далі будуємо модель ARIMA (1, 1,1) (рисунок 3.36):

ARIMA Model Results						
=====						
Dep. Variable:	D.y	No. Observations:	74			
Model:	ARIMA(1, 1, 1)	Log Likelihood	-217.585			
Method:	css-mle	S.D. of innovations	4.574			
Date:	Mon, 01 Jun 2020	AIC	443.171			
Time:	04:07:47	BIC	452.387			
Sample:	1	HQIC	446.847			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.8657	0.936	0.925	0.355	-0.969	2.701
ar.L1.D.y	0.5692	0.185	3.075	0.002	0.206	0.932
ma.L1.D.y	-0.2314	0.206	-1.123	0.262	-0.635	0.173
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	1.7569	+0.0000j	1.7569	0.0000		
MA.1	4.3220	+0.0000j	4.3220	0.0000		

R-square: 0.93649						
Durbin-Watson: 2.032						

Рисунок 3.36 – Результати оцінки моделі ARIMA(1, 1, 1)

У даному випадку характеристики адекватності моделі дещо кращі, порівняно з попередньою: $R^2 = 0.93649$, $DW = 2.032$.

Інформаційний критерій Акайке зменшився $AIC=443.171$, що також вказує на перевагу даної моделі.

Характеристики якості однокрокового прогнозу наступні (рисунок 3.37).

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.4775389	4.538603	3.080019	0.4241345	3.7954	0.9109104

Рисунок 3.37 – Характеристики якості однокрокового прогнозу ARIMA(1,1,1)

Можемо зробити висновок, що у порівнянні з попередньою моделлю, дана модель краще підходить для прогнозування.

Будуємо модель ARIMA(1,1,0) (рисунок 3.38), а також однокроковий прогноз та знаходимо його характеристики (рисунок 3.39).

ARIMA Model Results						
=====						
Dep. Variable:	D.y	No. Observations:	74			
Model:	ARIMA(1, 1, 0)	Log Likelihood	-218.127			
Method:	css-mle	S.D. of innovations	4.608			
Date:	Mon, 01 Jun 2020	AIC	442.355			
Time:	13:29:32	BIC	449.167			
Sample:	1	HQIC	445.012			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.8858	0.837	1.059	0.290	-0.754	2.526
ar.L1.D.y	0.3647	0.107	3.403	0.001	0.155	0.575
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	2.7423	+0.0000j	2.7423	0.0000		

R-square: 0.93769						
Durbin-Watson: 2.054						

Рисунок 3.38 – Результати оцінки моделі ARIMA(1, 1, 0)

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.5474143	4.579973	3.060681	0.4646188	3.693701	0.9140947

Рисунок 3.39 – Характеристики якості однокрокового прогнозу ARIMA(1,1,0)

Будуємо модель ARIMA(0,1,1) (рисунок 3.40), а також однокроковий прогноз та знаходимо його характеристики (рисунок 3.41):

ARIMA Model Results						
=====						
Dep. Variable:	D.y	No. Observations:	74			
Model:	ARIMA(0, 1, 1)	Log Likelihood	-219.812			
Method:	css-mle	S.D. of innovations	4.716			
Date:	Mon, 01 Jun 2020	AIC	445.625			
Time:	04:12:51	BIC	452.537			
Sample:	1	HQIC	448.382			
=====						
	coef	std err	z	P> z	[0.025	0.975]
const	0.8902	0.687	1.295	0.195	-0.457	2.238
ma.L1.D.y	0.2573	0.090	2.872	0.004	0.082	0.433
=====						
Roots						
	Real	Imaginary	Modulus	Frequency		
MA.1	-3.8868	+0.0000j	3.8868	0.5000		
=====						
R-square: 0.93763						
Durbin-Watson: 1.880						

Рисунок 3.40 – Результати оцінки моделі ARIMA(0, 1, 1)

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.6938812	4.705995	3.131173	0.5568392	3.755416	0.9351476

Рисунок 3.41 – Характеристики якості однокрокового прогнозу ARIMA(0,1,1)

Будуємо модель ARIMA(2,1, 0) (рисунок 3.42), а також однокроковий прогноз та знаходимо його характеристики (рисунок 3.43):

ARIMA Model Results						
=====						
Dep. Variable:	D.y	No. Observations:	74			
Model:	ARIMA(2, 1, 0)	Log Likelihood	-217.165			
Method:	css-mle	S.D. of innovations	4.547			
Date:	Mon, 01 Jun 2020	AIC	442.330			
Time:	04:10:25	BIC	451.546			
Sample:	1	HQIC	446.006			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.8570	0.972	0.882	0.378	-1.048	2.762
ar.L1.D.y	0.3041	0.114	2.667	0.008	0.081	0.528
ar.L2.D.y	0.1604	0.115	1.398	0.162	-0.065	0.385
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	1.7227	+0.0000j	1.7227	0.0000		
AR.2	-3.6180	+0.0000j	3.6180	0.5000		

R-square: 0.99573						
Durbin-Watson: 1.967						

Рисунок 3.42 – Результати оцінки моделі ARIMA(2, 1, 0)

Training set error measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.4645182	4.511708	3.079421	0.4190366	3.733893	0.9196915

Рисунок 3.43 – Характеристики якості однокрокового прогнозу ARIMA(2,1,0)

Побудуємо модель множинної регресії, та порівняємо її характеристики з попередніми моделями (рисунок 3.44).

OLS Regression Results						
=====						
Dep. Variable:	LiquidityExp	R-squared:	0.939			
Model:	OLS	Adj. R-squared:	0.935			
Method:	Least Squares	F-statistic:	271.1			
Date:	Mon, 01 Jun 2020	Prob (F-statistic):	3.37e-42			
Time:	13:13:17	Log-Likelihood:	-250.46			
No. Observations:	76	AIC:	510.9			
Df Residuals:	71	BIC:	522.6			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	73.1772	1.597	45.820	0.000	69.993	76.362
Liabilities	9.824e-06	0.000	0.063	0.950	-0.000	0.000
Deposit	0.0007	0.000	3.138	0.002	0.000	0.001
Loan_portfolio	-0.0003	5.91e-05	-4.720	0.000	-0.000	-0.000
Assets	-0.0001	0.000	-1.162	0.249	-0.000	0.000
=====						
Omnibus:	0.230	Durbin-Watson:	1.601			
Prob(Omnibus):	0.891	Jarque-Bera (JB):	0.014			
Skew:	0.012	Prob(JB):	0.993			
Kurtosis:	3.062	Cond. No.	5.10e+05			
=====						

Рисунок 3.44 –Результати оцінки моделі OLS

Занесемо всі отримані результати у таблицю 3.3.

Таблиця 3.3 – Характеристики побудованих моделей

Model	MAE	MAPE	RMSE	R ²	DW	AIC
ARIMA(0,1, 0)	3.305246	3.94246	4.971936	0.9321	1.738	450.963
ARIMA(1,1, 0)	3.060681	3.693701	4.579973	0.93679	2.054	442.355
ARIMA(0,1, 1)	3.131173	3.755416	4.705995	0.93763	1.880	445.625
ARIMA(2,1, 0)	3.079421	3.733893	4.511708	0.99573	1.967	442.330
ARIMA(1,1, 1)	3.080019	3.7954	4.538603	0.93649	2.032	443.117
OLS	4.83931	6.31223	6.53135	0.939	1.601	510.9

Порівнявши показники усіх моделей, для побудови прогнозу обираємо модель ARIMA(2,1, 0). Спрогнозуємо процент поточної ліквідності на 8 кроків вперед, тобто на чотири квартали 2020-го року та на чотири квартали 2021-го року (рисунок 3.43).

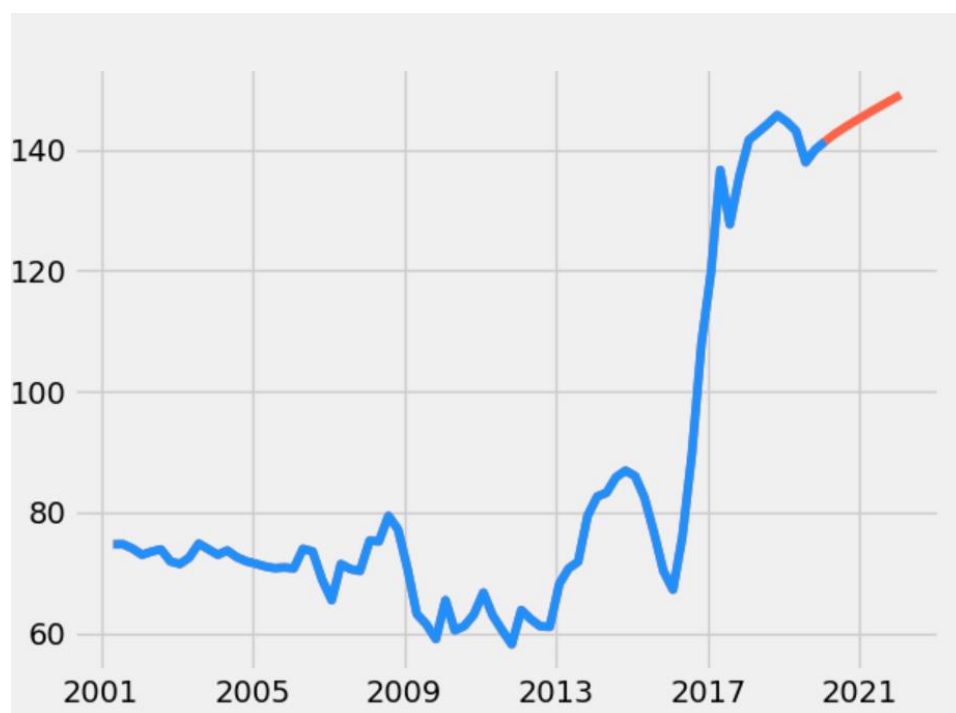


Рисунок 3.43 – Результати прогнозування ряду Н5 на 8 кроків вперед
Результати прогнозування наведемо у таблиці 3.4.

Таблиця 3.4 – Результати прогнозування ряду Н5

Дата	Значення прогнозу
2020q1	142.54498188
2020q2	143.59111933
2020q3	144.57359999
2020q4	145.51216305
2021q1	146.42730944
2021q2	147.32840502
2021q3	148.22152665
2021q4	149.11000128

Загалом, згідно прогнозу, можемо зробити висновок, що процент поточної ліквідності продовжить зростати, та значно перевищує норматив 40%, що є гарним показником фінансового стану банку. Адже це значить, що сума оборотних активів значно перевищує суму поточних зобов'язань.

3.3 Висновки до розділу 3

У першій частині розділу було описано вхідні дані та побудовано графіки досліджуваних часових рядів.

У другій частині даного розділу було виконано заміну пропущених значень даних, методом випадкового лісу, та були наведені результати даного процесу. Далі було виконано процедуру фільтрації, щоб позбутися шуму та зайвої інформації. Також було проведено тест Діккі-Фулера, за допомогою якого ЧР перевірялись на стаціонарність. Крім того, було проведено низку

експериментів та побудовано різноманітні математичні моделі, серед яких було обрано кращу для прогнозування.

А також було спрогнозовано процент кредитного ризику та ризику ліквідності АТ КБ «ПРИВАТБАНК» на наступні два роки.

РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

4.1. Постановка задачі проектування

Функціонально – вартісний аналіз програми для прогнозування та аналізу кредитних ризиків за допомогою регресійних моделей прогнозування часових рядів.

Основною задачею проектування є створення програмного продукту для аналізу та прогнозування банківських ризиків, за допомогою регресійних моделей прогнозування часових рядів.

Вимоги до продукту:

- а) ПП має бути простий у користуванні;
- б) ПП повинен функціонувати на різних ПК;
- в) ПП повинен мати високу швидкість роботи;
- г) ПП має коректно працювати з різними вибірками;

4.2. Обґрунтування функцій та параметрів програмного продукту

Виходячи з конкретних цілей, які реалізуються :

F1 – обрання мови програмування:

- а) об'єктно-орієнтована мова програмування високого рівня Python,
- б) мова програмування і програмне середовище R.

F2 – реалізація алгоритмів програмного продукту:

- а) використання існуючих модулів та бібліотек,
- б) самостійне створення та написання алгоритмів.

F3 – створення інтерфейсу програмного продукту:

- а) графічний інтерфейс користувача,

б) консольний інтерфейс користувача.

З огляду на вищенаведені варіанти основних функцій побудуємо морфологічну карту (рисунок 4.1).

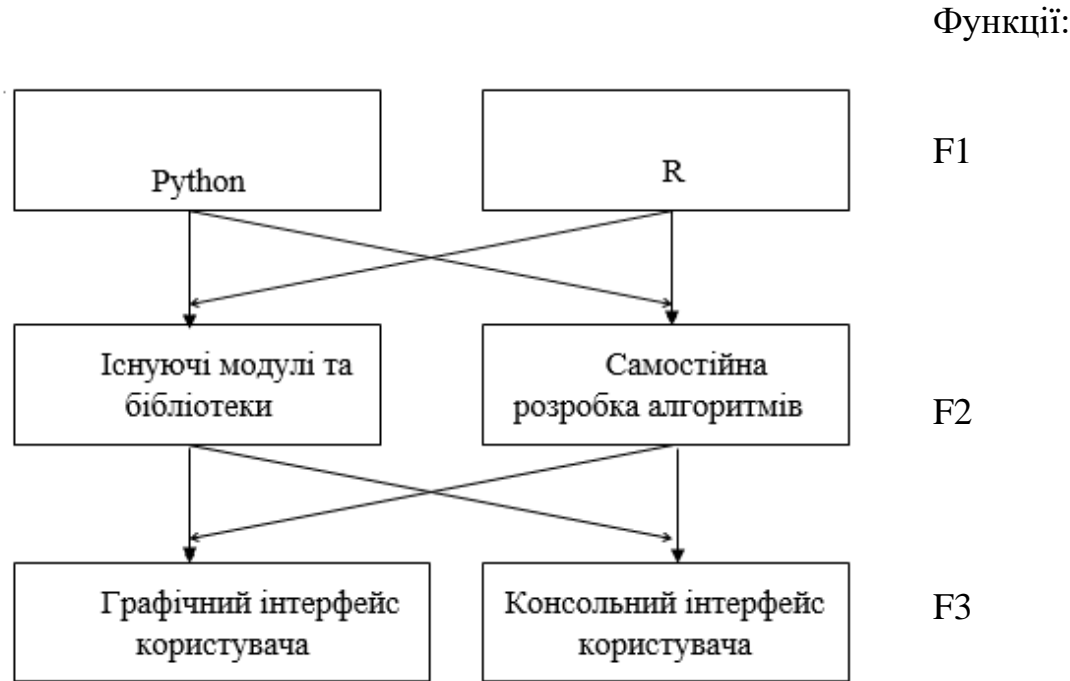


Рисунок 4.1 – Морфологічна карта

На морфологічній карті зображено усі можливі комбінації реалізації ПП.

З огляду на морфологічну карту будуємо позитивно-негативну матрицю варіантів (таблиця 4.1).

Таблиця 4.1 – Позитивно-негативна матриця

Основні функції	Варіанти реалізації	Переваги	Недоліки
F1	А	Легкість реалізації завдань та цілей поставленої задачі	Повільна швидкість виконання операцій
	Б	Легкість реалізації завдань та цілей поставленої задачі	Складність мови

Продовження таблиці 4.1

Основні функції	Варіанти реалізації	Переваги	Недоліки
F2	А	Витрачається значно менше часу при написанні коду програми, зручніше та простіше орієнтуватися у написаному коді	Алгоритми менш орієнтовані на конкретну задачу
	Б	Створені алгоритми більш орієнтовані на поставлену задачу	Складна реалізація, значна затрата часу
F3	А	Більш наглядний та привабливий з візуальної точки зору	Складність у реалізація, та налаштуванні. Затрата часу
	Б	Легкість реалізації	Дещо складний у користуванні пересічним користувачем

Проаналізувавши позитивно-негативну таблицю було обрано наступні такі варіанти реалізації програмного продукту:

$F_1 \text{ А} - F_2 \text{ А} - F_3 \text{ Б}$

$F_1 \text{ А} - F_2 \text{ Б} - F_3 \text{ Б}$

Щоб оцінити прототип програмного продукту використаємо наступні параметри.

Функція F_1 залежить від усіх трьох параметрів, а саме від часу, витраченого на реалізації обраних моделей, швидкості обробки вибірок та запитів програмою і також від кількості рядків коду у готовому програмному продукті (X_1, X_2, X_3).

Функція F_2 залежить від двох параметрів X_1 та X_3 , тобто від часу, витраченого на реалізації обраних моделей, та від кількості рядків коду у готовому програмному продукті.

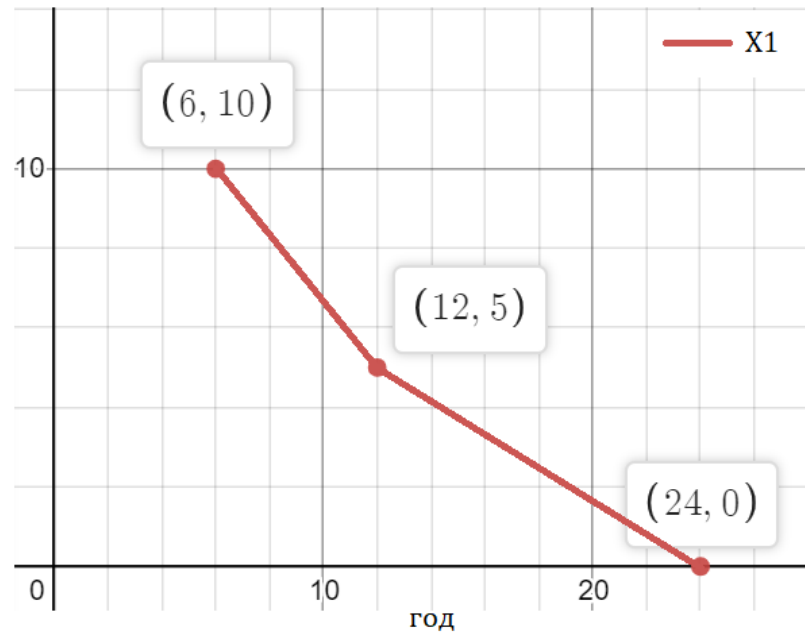
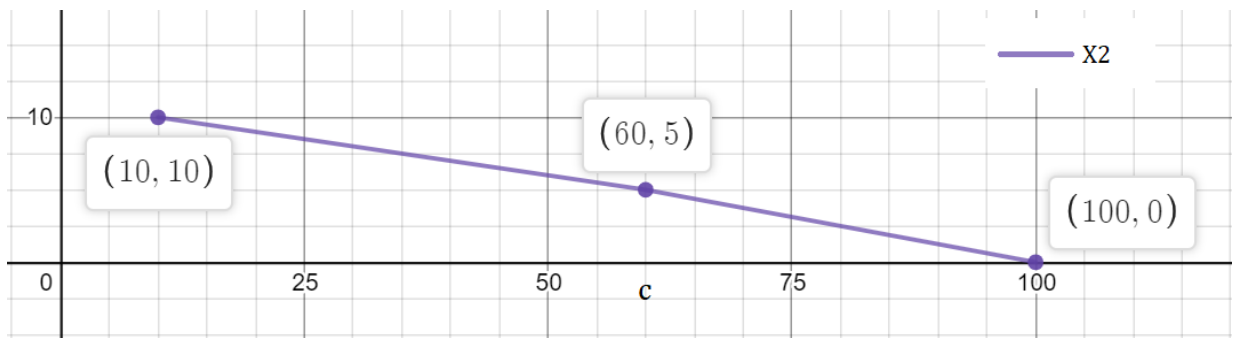
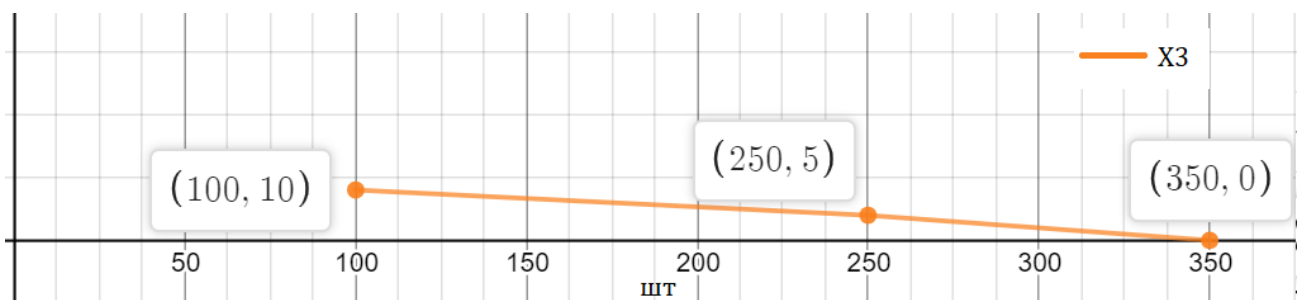
Функція F_3 залежить від параметру X_3 : кількості рядків коду у готовому програмному продукті.

Будуємо таблицю 4.2 з параметрами.

Таблиця – 4.2 Система параметрів програмного продукту

Назва параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Час, витрачений на реалізацію обраних моделей	X_1	год	24	12	6
Швидкість обробки вибірок та запитів програмою	X_2	с	100	60	10
Кількість рядків коду у готовому програмному продукті	X_3	шт	350	250	150

Використовуючи дані таблиці будуємо графіки (рисунок 4.2, рисунок 4.3, рисунок 4.4).

Рисунок 4.2 – Значення параметру X_1 Рисунок 4.3 – Значення параметру X_2 Рисунок 4.4 – Значення параметру X_3

Вагомість параметрів будемо оцінювати за допомогою методів попарного зрівняння. Так як параметра усього три, то ранги варіюються від 1 до 3. де 3 – це найвищий ранг. Результати продемонстровано у таблицях 4.3-4.4.

Таблиця 4.3 – Результат оцінки параметрів

Параметр	Ранг параметру по оцінці експерта							Сума рангів, R_i	Відхилення Δ_i	Квадрат відхилення, $(\Delta_i)^2$
	1	2	3	4	5	6	7			
X1	3	3	2	3	3	3	3	20	6	36
X2	2	2	3	2	2	1	2	14	0	0
X3	1	1	1	1	1	2	1	8	-6	36
Разом	6	6	6	6	6	6	6	42	0	72

Таблиця 4.4 – Попарне зрівняння параметрів

Параметр и	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 та X2	>	>	<	>	>	>	>	>	1.5
X1 та X3	>	>	>	>	>	>	>	>	1.5
X2 та X3	>	>	>	>	>	<	>	>	1.5

Визначемо коефіцієнт узгодженості:

$$W = \frac{12 * 72}{49(27 - 3)} \approx 0.735 > 0.67$$

Оскільки коефіцієнт узгодженості більше нормативного, результати можемо вважати достовірними.

Використаємо результати попарного порівняння та обчислимо вагомість кожного з параметрів. Результати наведемо у таблиці 4.5.

Таблиця 4.5 – Розрахунок вагомості параметрів

X _i	Параметри X _j			Перший крок		Другий крок		Третій крок	
	X ₁	X ₂	X ₃	B _i	K _{Bi}	B' _i	K' _{Bi}	B'' _i	K'' _{Bi}
X ₁	1.0	1.5	1.5	4	0.445	11.5	0.46	31.75	0.461
X ₂	0.5	1.0	1.5	3	0.333	8	0.32	21.85	0.317
X ₃	0.5	0.5	1.0	2	0.222	5.5	0.22	15.25	0.222
Усього:				9	1	25	1	68.85	1

Обчислимо коефіцієнту якості. Результати наведемо у таблиці 4.6.

Таблиця 4.6 – Розрахунок коефіцієнту якості

Основна функція	Варіант реалізації	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт якості
F ₁	А	X ₁	11	5.84	0.461	2.692
		X ₂	15	9.5	0.317	3.011
		X ₃	220	6	0.222	1.332
F ₂	А	X ₁	8	8.33	0.461	3.84
		X ₃	100	10	0.317	3.17
F ₂	Б	X ₁	15	3.75	0.461	1.728
		X ₃	300	2.5	0.222	0.555
F ₃	Б	X ₃	200	6.67	0.222	1.48

$$K_1 = 2.692 + 3.011 + 1.332 + 3.84 + 3.17 + 1.48 = 15.525$$

$$K_2 = 2.692 + 3.011 + 1.332 + 1.728 + 0.555 + 1.48 = 10.798$$

Так як варіант №1 має більший коефіцієнт якості, то він є кращим.

4.3 Економічний аналіз варіантів розробки

Щоб оцінити трудомісткості розробки проведемо певні розрахунки. Варіанти реалізації продукту можна описати наступним чином:

- 1) Розробка загальної структури ПП
- 2) Створення ПП
 - А) Ознайомлення та використання готових модулів і бібліотек до обраних даних
 - Б) Розробка власних алгоритмів обраних моделей
- 3) Демонстрація та візуалізація результатів

Для першого завдання маємо: алгоритм групи складності 3, ступінь новизни В, вид використаної інформації ПП

$$T_p = 12, K_{\Pi} = 0.6, K_{CK} = 1.07, K_{CT.M} = 1.2,$$

$$T_1 = 12 \cdot 0.6 \cdot 1.07 \cdot 1.2 = 9,244 \text{ (людино-днів)}$$

Для другого завдання (А) маємо: алгоритм групи складності 2, ступінь новизни В, вид використаної інформації ПП

$$T_p = 19, K_{\Pi} = 1.2, K_{CK} = 0.8, K_{CT.M} = 1.3,$$

$$T_{2a} = 19 \cdot 1.2 \cdot 0.8 \cdot 1.3 = 23,712 \text{ (людино-днів)}$$

Для другого завдання (Б) маємо: алгоритм групи складності 1, ступінь новизни Б, вид використаної інформації ПП

$$T_p = 64, K_{\Pi} = 2.02, K_{CK} = 1.08, K_{CT.M} = 1.5,$$

$$T_{26} = 64 \cdot 2.02 \cdot 1.08 \cdot 1.5 = 209,433 \text{ (людино-днів)}$$

Для першого завдання маємо: алгоритм групи складності 2, ступінь новизни Г, вид використаної інформації БД

$$T_p = 12, K_n = 0.36, K_{ск} = 1.07, K_{ст.м} = 1.3,$$

$$T_3 = 12 \cdot 0.36 \cdot 1.07 \cdot 1.3 = 6,009 \text{ (людино-днів)}$$

Тоді, у випадку варіанта А, маємо:

$$T = (9,244 + 23,712 + 6,009) \cdot 8 = 311.72 \text{ (людино-годин)}$$

У випадку варіанта Б, маємо:

$$T = (9,244 + 209,433 + 6,009) \cdot 8 = 1797.488 \text{ (людино-годин)}$$

В розробці бере участь один програміст з окладом 15500 грн та аналітик з окладом 22000 грн.

Розрахуємо середню заробітну плату за годину:

$$C_q = \frac{15500 + 22000}{2 * 22 * 8} = 106,53$$

Тоді заробітна плата для кожного з варіантів реалізації(грн):

$$C_{зп1} = 106,53 * 311.72 = 33207,53,$$

$$C_{зп2} = 106,53 * 1797.488 = 191486,39$$

Відрахування ЄСВ (22%)(грн):

$$C_{від1} = 33207,53 * 0,22 = 7307.65,$$

$$C_{від2} = 191486,39 * 0,22 = 42127,00$$

Далі визначаємо витрати на оплату однієї машино-години. Враховуючи, що заробітна плата одного програміста з окладом 15500 грн з коефіцієнтом зайнятості 0,4.

$$C_{\Gamma} = 12 * 15500 * 0,4 = 74400 \text{ грн}$$

Враховуючи додаткову заробітну плату:

$$C_{ЗП} = 74400 * (1 + 0,3) = 96720 \text{ грн}$$

Відрахування на соціальне страхування 22%

$$C_{ВІД} = 96720 * 0,22 = 21278,4 \text{ грн}$$

Розрахуємо амортизаційні підрахунки (амортизація 25%, вартість ЕОМ 22000 грн)

$$C_A = K_{TM} * K_A * C_{ПР} = 1,15 * 0,25 * 22000 = 6325 \text{ грн}$$

Розрахуємо витрати на ремонт та профілактику:

$$C_P = K_{TM} * C_{ПР} * K_P = 1,15 * 22000 * 0,05 = 1265 \text{ грн}$$

Розрахуємо ефективний годинний фонд часу ПК за рік:

$$T_{ЕФ} = (365 - 104 - 11 - 8) * 8 * 0,9 = 1742,4 \text{ год}$$

Розрахуємо витрати на електроенергію

$$C_{\text{ЕЛ}} = 1742,4 * 0,29 * 0,927 * 1,75 = 819.72 \text{ грн}$$

Накладні витрати рівні:

$$C_{\text{Н}} = 22000 * 0,67 = 14740 \text{ грн.}$$

Отже експлуатаційні витрати(грн):

$$C_{\text{ЕКС}} = 96720 + 21278.4 + 6325 + 1265 + 819.72 + 14740 = 141148.12$$

Тоді собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = \frac{141148.12}{1742,4} = 81,00 \text{ грн/год}$$

Враховуючи, що всі роботи ведуться на ЕОМ, розраховуємо витрати на оплату машинного часу для кожного варіанту:

$$C_{\text{М}} = 81,00 * 311.72 = 25249,32 \text{ грн,}$$

$$C_{\text{М}} = 81,00 * 1797.488 = 145596,528 \text{ грн}$$

Накладні витрати відповідно

$$C_{\text{Н}} = 33207,53 * 0,67 = 22249,045 \text{ грн,}$$

$$C_{\text{Н}} = 191486,39 * 0,67 = 128295,881 \text{ грн}$$

Розраховуємо повну вартість розробки в залежності від варіантів:

$$C_{\text{ПП}} = 33207,53 + 7307.65 + 25249,32 + 22249,045 = 88023,545,$$

$$\begin{aligned} C_{\text{ПП}} &= 191486,39 + 42127,00 + 145596,528 + 128295,881 \\ &= 507505,799 \end{aligned}$$

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$\begin{aligned} K_{\text{ТЕР}j} &= K_{\text{К}j} / C_{\text{Ф}j}, \\ K_{\text{ТЕР}1} &= \frac{15.525}{88023,545} = 17,63 * 10^{-5}, \\ K_{\text{ТЕР}2} &= \frac{10.798}{507505,799} = 2,12 * 10^{-5} \end{aligned}$$

4.4 Висновки до розділу 4

Отже, проаналізувавши отримані результати, можемо зробити висновок, що більш ефективним буде перший варіант з коефіцієнтом техніко-економічного рівня $K_{\text{ТЕР}1} = 17,63 * 10^{-5}$. Тож, після проведеного функціонально – вартісного аналізу приймаємо рішення реалізовувати перший варіант, який передбачає використання готових модулів та бібліотек обраної мови програмування Python та подання результатів роботи ПП у вигляді консольного інтерфейсу.

ВИСНОВКИ

У першому розділі дипломної роботи було розглянуто сутність поняття «кредитний ризик», та його класифікацію. Також було наведено три способи оцінки кредитного ризику та декілька підходів до його прогнозування. Крім того була складена порівняльна таблиця існуючих комп'ютерних систем для побудови моделей фінансово-економічних процесів, з якої можна зробити висновок, що на сьогодні найзручнішим ПЗ а також мовою програмування – є мова R.

У другому розділі були розглянуті основні поняття та характеристики часового ряду. Наведені статистичні тести, що дозволяють провести перевірку часового ряду на стаціонарність. Також були розглянуті методи до заповнення пропусків – метод випадкового лісу, метод k-найближчих сусідів, та підхід до фільтрації даних у вигляді експоненціального фільтру. Також було наведено чотири математичні моделі для опису та прогнозування ЧР. Крім того було наведено критерії, що дозволяють оцінити адекватність відповідної математичної моделі та якість прогнозу.

У першій частині третього розділу було описано вхідні дані та побудовано графіки досліджуваних часових рядів. Було виконано заміну пропущених значень даних, методом випадкового лісу, та були наведені результати даного процесу. Далі було виконано процедуру фільтрації, щоб позбутися шуму та зайвої інформації. Також було проведено тест Діккі-Фулера, за допомогою якого ЧР перевірялись на стаціонарність. Крім того, було проведено низку експериментів та побудовано різноманітні математичні моделі, серед яких було обрано кращу для прогнозування.

Також було спрогнозовано процент кредитного ризику та процент поточної ліквідності АТ КБ «ПРИВАТБАНК» на наступні два роки. Загалом, згідно прогнозу, можемо зробити висновок, що процент кредитного ризику на

одного контрагента продовжує зростати, проте не перевищує норму - 25%, а процент поточної ліквідності продовжує зростати, та значно перевищує граничне значення 40%, що є гарним показником фінансової стабільності банку. Адже це значить, що сума оборотних активів значно перевищує суму поточних зобов'язань.

ЛІТЕРАТУРА

1. Балджи М.Д., Карпов В.А., Ковальов А.І., Костусєв О.О., Котова І.М., Сментина Н.В. Обґрунтування господарських рішень та оцінка ризиків: навч. посіб. / Одеса: ОНЕУ, 2013. 670 с.
2. Клапків М.С. Страхування фінансових ризиків: практ. посіб. Тернопіль: Економічна думка & Карт-бланш, 2002. 570 с.
3. Шумелда Я. С. Навчальний посібник для студентів економічних спеціальностей: навч. посіб. Тернопіль: Джура, 2004. 280 с.
4. Горский М.А. Проблематика коэффициентного подхода к оценке совокупного риска кредитного портфеля коммерческого банка. *Международный журнал экспериментального образования*. 2016. № 5. С. 109-114
5. Бідюк П.І. Часові ряди: моделювання та прогнозування / Київ: ЕКМО, 2004. 144 с.
6. ДСТУ 3008:2015. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення. [На заміну ДСТУ 3008-95; чинний від 2015-06-22]. Вид. офіц. Київ: ДП «УкрНДНЦ», 2016. 31 с. (Інформація та документація).
7. Бідюк П.І., Коршевніук Л.О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень: навч. посіб. / ННК «Інститут прикладного системного аналізу» Національний технічний університет України «Київський політехнічний інститут», 2010. 340 с.

ДОДАТОК А

Ілюстративний матеріал

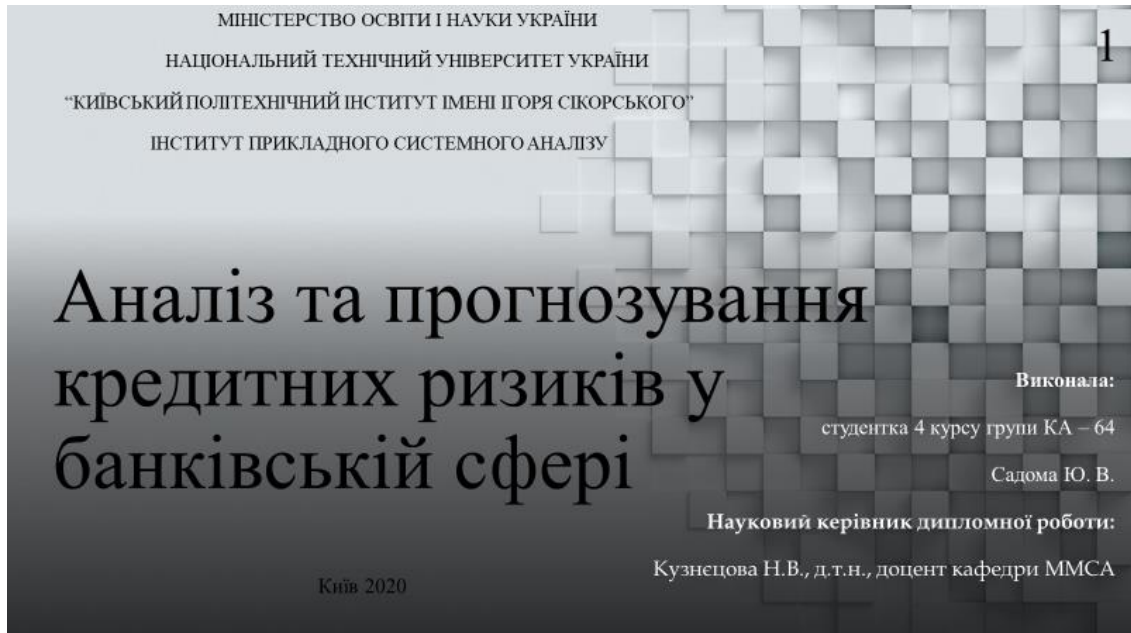


Рисунок 1 – Початок презентації

2

Об'єкт, предмет та мета дослідження

- **Об'єкт дослідження** – фінансові показники АТ КБ «ПРИВАТБАНК» з 2001-го року по 2019-го рік, а саме дані про процент кредитного ризику та процент поточної ліквідності.
- **Предмет дослідження** – методи прогнозування часових рядів, а також критерії адекватності математичних моделей і якості прогнозів.
- **Метою даної дипломної роботи** є створення та реалізація програмного продукту для аналізу та прогнозування кредитних та інших фінансових ризиків, що впливають на стабільність банку. Створений ПЗ може спростити деякі аспекти роботи та бути корисним у банківській сфері.

Рисунок 2 – Об'єкт, предмет і мета дослідження

Завдання дипломної роботи

- А) Аналіз існуючих методів та моделей для прогнозування часових рядів
- Б) Пошук даних та формування вибірки
- В) Вибір та реалізація методів для попередньої обробки даних
- Г) Вибір та реалізація методів та моделей для прогнозування часових рядів
- Д) Створення програмного продукту
- Е) Застосування програмного продукту на реальних даних, аналіз результатів

Рисунок 3 – Завдання дипломної роботи

Основні поняття

- **Ризиком** називають деяку невизначену подію чи умову, яка може внести як позитивний так і негативний вплив у роботу та розвиток певної компанії або підприємства, та може призвести до грошових придбань чи втрат.
- **Кредитний ризик або ж ризик дефолту** – це ризик, що з'являється при своєчасному невиконанні учасником-контрагентом (боржником) фінансових зобов'язань, тобто позичальник не відшкодовує свої боргові виплати, що у свою чергу призводить до фінансових втрат банку.
- **Ризик ліквідності** - фінансовий ризик, пов'язаний з неспроможністю розрахуватися за зобов'язаннями в належні строки без істотних збитків.

Рисунок 4 – Основні поняття

5

Використані моделі

- Модель авторегресії AR(p) $y_t = a_0 + a_1y_{t-1} + a_2y_{t-2} + \dots + a_py_{t-p} + \varepsilon_t$
- Модель авторегресії з ковзним середнім ARMA(p,q)

$$y_t = a_0 + a_1y_{t-1} + a_2y_{t-2} + \dots + a_py_{t-p} + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}$$
- Інтегрована модель авторегресії з ковзним середнім ARIMA(p,d,q)

$$\Delta^d y_t = a_0 + a_1\Delta y_{t-1} + a_2\Delta y_{t-2} + \dots + a_p\Delta y_{t-p} + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q},$$
- Модель множинної лінійної регресії $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p + \varepsilon$

Рисунок 5 – Використані моделі

6

Критерії адекватності математичних моделей і якості прогнозів

AIC	інформаційний критерій Акайке	$2k - 2\ln(L)$
R^2	коефіцієнт детермінації	$1 - \frac{V(y x)}{V(y)}$
DW	критерій Дарбіна-Уотсона	$\frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$
RMSE	корінь з середньоквадратичної похибки	$\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$
MAPE	середня абсолютна похибка в процентах	$\frac{1}{N} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{ y_i } \times 100\%$
MAE	середня абсолютна похибка	$\frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i $

Рисунок 6 – Критерії адекватності математичних моделей і якості прогнозів



Рисунок 7 – Алгоритм функціонування програмного продукту

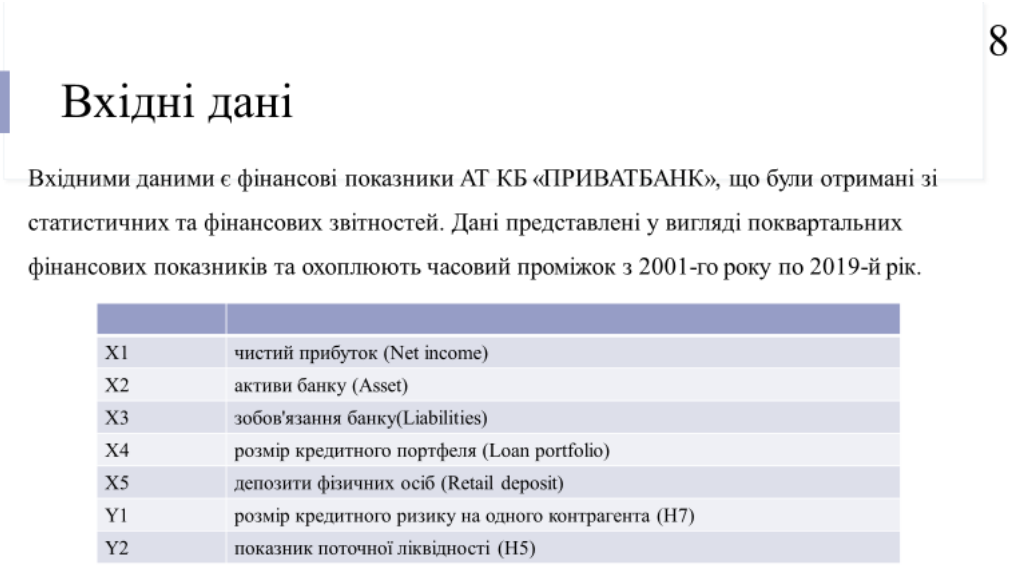


Рисунок 8 – Вхідні дані

Заповнення пропусків даних методом випадкового лісу

A	B	C	D	E	F	G	H	I
1	Date	Net Income	Assets	Liabilities	Loan_portfolio	Retail deg H7	H5	
2	2001q1	7	5742	5460	1995	898	24.6072	-1
3	2001q2	15	6271	5980	2442	1061	20.5312	-1
4	2001q3	30	7627	7316	2828	1200	18.5771	72.3565
5	2001q4	34	8254	7890	3365	1408	20.5771	70.5247
6	2002q1	8	8036	7664	3366	1716	20.6072	74.7045
7	2002q2	19	9090	8690	3958	1954	20.5312	74.8235
8	2002q3	30	10248	9837	4386	2317	21.5374	67.21
9	2002q4	148	10858	10311	4437	2615	19.6052	70.4715
10	2003q1	7	11551	10996	5279	3123	20.6072	74.9776
11	2003q2	21	13824	13255	5951	3380	17.5154	80.22
12	2003q3	34	16784	16202	7019	3800	20.5771	71.7654
13	2003q4	60	17490	16534	7502	4287	22.6196	70.8145
14	2004q1	14	17874	16904	8611	4967	19.5312	75.3968
15	2004q2	143	21676	20577	11010	5422	-1	-1
16	2004q3	236	15589	14397	11249	6170	-1	-1
17	2004q4	163	14668	13204	11109	5558	-1	-1
18	2005q1	205	16469	-1	12715	6585	20.562	70.0342
19	2005q2	215	17808	15916	13724	7091	23.4939	70.0948

Рисунок 9 – Заповнення пропусків даних методом випадкового лісу

Результати застосування експоненціального фільтру для ряду Y1 розміру кредитного ризику на одного контрагента

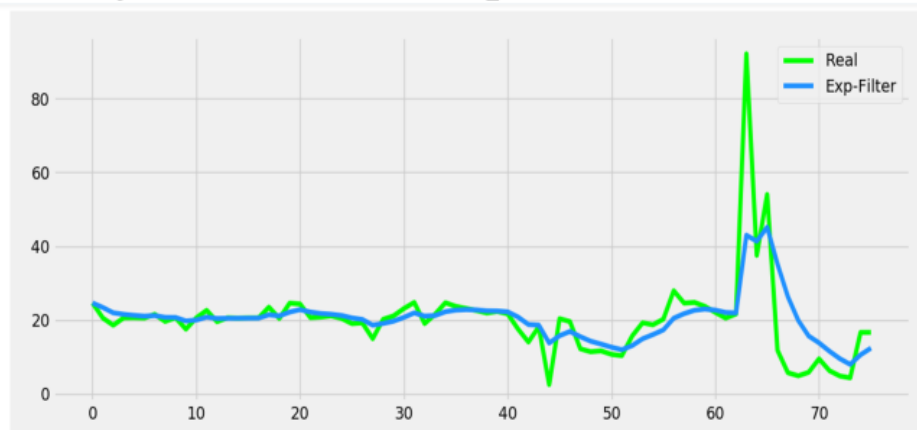


Рисунок 10 – Результати застосування експоненціального фільтру для ряду Y1 розміру кредитного ризику на одного контрагента

Результати застосування експоненціального 11
фільтру для ряду Y2 показника поточної
ліквідності

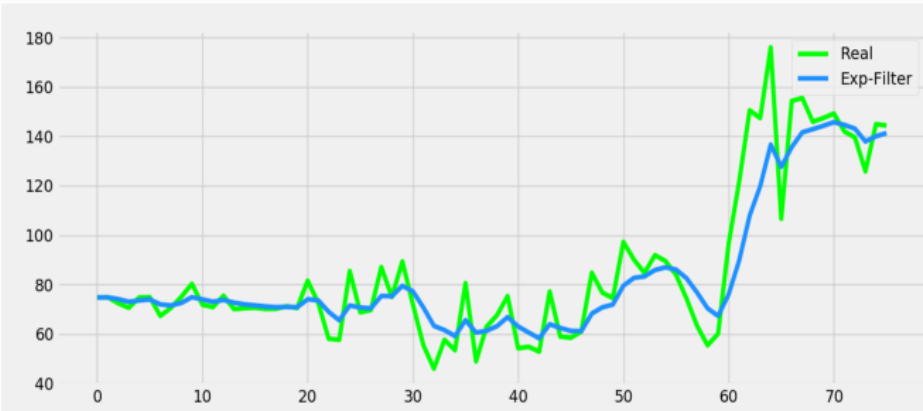


Рисунок 11 – Результати застосування експоненціального фільтру для ряду Y2 показника поточної ліквідності

Характеристики побудованих моделей для 12
ряду Y1 розміру кредитного ризику на
одного контрагента

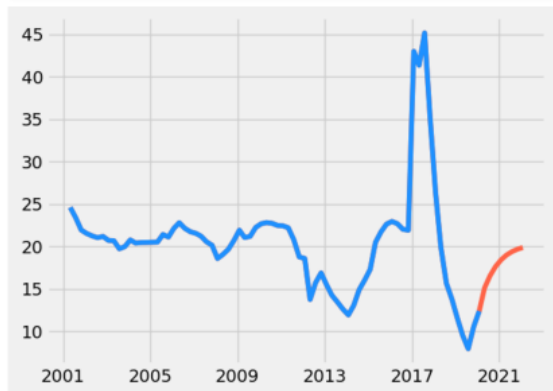
$p - value = 0.002$

Model	MAE	MAPE	RMSE	R ²	DW	AIC
AR(1)	1.497191	7.434088	3.156664	0.853560	1.711	393.59
AR(2)	1.388911	6.76876	3.055489	0.752142	1.991	394.93
ARMA(1,2)	1.238694	6.414319	2.83752	0.98564	1.994	382.20
ARMA(2,1)	1.508952	7.218115	2.875344	0.847	1.694	389.086
OLS	2.42254	13.32067	3.924754	0.603	1.497	435.5

Рисунок 12 –Характеристики побудованих моделей для ряду Y1 розміру кредитного ризику на одного контрагента

Результати прогнозування проценту кредитного ризику на одного контрагента на основі моделі ARMA(1,2)

13



Дата	Значення прогнозу
2020q1	15.12625538
2020q2	16.55940161
2020q3	17.64218492
2020q4	18.42119572
2021q1	18.98155416
2021q2	19.38457078
2021q3	19.67462327
2021q4	19.88344807

Рисунок 13 – Результати прогнозування

Характеристики побудованих моделей для ряду Y2 (показника поточної ліквідності)

14

 $p - value = 0.87 \longrightarrow p - value = 9.427e^{-23}$

Model	MAE	MAPE	RMSE	R ²	DW	AIC
ARIMA(0,1,0)	3.305246	3.94246	4.971936	0.9321	1.738	450.963
ARIMA(1,1,0)	3.060681	3.693701	4.579973	0.93679	2.054	442.355
ARIMA(0,1,1)	3.131173	3.755416	4.705995	0.93763	1.880	445.625
ARIMA(2,1,0)	3.079421	3.733893	4.511708	0.99573	1.967	442.330
ARIMA(1,1,1)	3.080019	3.7954	4.538603	0.93649	2.032	443.117
OLS	4.83931	6.31223	6.53135	0.939	1.601	510.9

Рисунок 14 – Характеристики побудованих моделей для ряду H5

Результати прогнозування показника поточної ліквідності на основі моделі ARIMA(2,1, 0)

15



Рисунок 15 – Результати прогнозування 2

Висновки

16

- 1) Зроблено аналіз основних методів та моделей для прогнозування часових рядів, на основі чого були обрані та реалізовані наступні моделі: авторегресії, авторегресії з ковзним середнім, інтегрованої авторегресії з ковзним середнім та множинної лінійної регресії.
- 2) У якості вхідних даних було обрано фінансові показники АТ КБ «ПРИВАТБАНК» з 2001-го року по 2019-й рік, що були отримані зі статистичних та фінансових звітностей. та охоплюють часовий проміжок з 2001-го року по 2019-го рік.
- 3) Для попередньої обробки даних обрано та реалізовано метод випадкового лісу а також експоненціальний фільтр.

Рисунок 16 – Висновки

17

Висновки

- 3) Створено програмний продукт мовою Python 3.7.0.
- 4) Спрогнозовано розмір кредитного ризику на одного контрагента та показник поточної ліквідності АТ КБ «ПРИВАТБАНК» на 4 квартали 2020-го та 2021-го року.
- 5) Згідно прогнозу, зроблено висновок, що процент кредитного ризику на одного контрагента продовжуватиме зростати, проте не перевищить норму у 25%, а процент поточної ліквідності продовжуватиме зростати, та значно перевищить граничне значення 40%, що є гарним показником, адже це вказує на те, що сума оборотних активів значно перевищує суму поточних зобов'язань. Загалом, отримані результати свідчать про фінансову стабільність банку.

Рисунок 17 – Висновки2

18

ДЯКУЮ ЗА УВАГУ!

Рисунок 18 – Завершальний слайд

ДОДАТОК Б

Лістинг програми

```
import numpy as np
import pandas as pd
from missingpy import KNNImputer
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from statsmodels.stats.stattools import durbin_watson
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import statsmodels.api as sm
from pandas import DataFrame
from missingpy import MissForest
plt.style.use('fivethirtyeight')
from pandas import read_csv
from pandas import datetime
from matplotlib import pyplot
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.arima_model import ARMA
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.iolib.table import SimpleTable
#Завантажуємо дані з пропусками
Data_Risks = pd.read_csv('Credit.csv')
X = list(Data_Risks.values)
tX = []
```

```

for i in X:
    tmp = []
    for j in i:
        if (j == -1):
            tmp += [np.nan]
        else:
            tmp += [float(j)]
    tX += [tmp]
X=tX

pd.options.display.max_rows = 1000
print(Data_Risks)

#Застосовуємо метод випадкового лісу
imputer = MissForest()
xMissForest = imputer.fit_transform(X)
for i in range(len(xMissForest)):
    for j in range(len(xMissForest[i])):
        xMissForest[i][j] = (float)(xMissForest[i][j])

Data_MissForest = pd.DataFrame(data =
xMissForest,columns=Data_Risks.columns)

Data_MissForest
print(Data_MissForest)

Data_MissForest.to_csv(r'MF.csv', index = False)

plt.figure(figsize=(15,5))
plt.plot(Data_Risks.index, Data_Risks['Lic'], color = 'lime')
plt.grid()
plt.show()

#Застосовуємо експоненціальний фільтр
def exponential_smoothing(series, theta):
    result = [series[0]]

```

```

    for i in range(1, len(series)):
        result.append(theta * series[i] + (1 - theta) * result[i-1])
    return result

Credit = pd.Series(Data_Risks['Liquidity'])
t= [exponential_smoothing(Credit,0.3)]
for x in t: print(x)
plt.figure(figsize=(15,5))
plt.plot(Credit,color = 'lime', label = 'Real')
plt.plot(exponential_smoothing(Credit,0.3),color = 'dodgerblue',label='Exp-Filter')
plt.legend(loc = 'best')
plt.show()
series = pd.read_csv('Credit.csv')
Data_Risks = pd.read_csv('Linear.csv')
def Summa(arr):
    add = 0
    for i in arr:
        add += i
    return add
def Theil(y_test,y_pred):
    return (np.sqrt(np.mean((y_test-
y_pred)**2)))/((np.sqrt(np.mean(y_test**2)))+(np.sqrt(np.mean(y_pred**2))))
def mean_absolute_percentage_error(y_test, y_pred):
    return np.mean(np.abs((y_test - y_pred) / y_test)) * 100
size = 0.75
X_with_missing_values = np.array(Data_Risks.values)
all = 0
y_with_missing_values = pd.Series(series['Liquidity'])
y = np.array(y_with_missing_values)
print(y)

```



```

X_train, X_test, y_train, y_test = train_test_split(X_with_missing_values, y,
train_size=size, shuffle=False)
print(X_train)
print(y_train)
print(X_test)
print(y_test)
df = DataFrame(Data_Risks,columns=['Assets','Liabilities','Deposit'])
df1=DataFrame(series,columns=['i','Date','Profit','Assets','Liabilities','Loan_portfoli
o','Deposit','Equity','Credit_risk','Liquidity','Cexp','LiquidityExp'])
X = df1[['Liabilities','Deposit','Loan_portfolio','Assets']] # here we have 2 variables
for the multiple linear regression. If you just want to use one variable for simple
linear regression, then use X = df['Interest_Rate'] for example
Y = df1['LiquidityExp']
X = sm.add_constant(X) # adding a constant
#Множинна лінійна регресія
model = sm.OLS(Y, X).fit()
predictions = model.predict(X)
print_model = model.summary()
print(print_model)
DAT = pd.DataFrame({"Actual": Y,"Predict":predictions})
print("R-square:",round(abs(r2_score(Y,predictions)), 5))
print("Durbin-Watson:",round(durbin_watson(predictions-Y), 5))
print("Sum-Square:",round(np.sum((predictions-Y)**2), 5))
print("MSE:",round(mean_squared_error(Y,predictions), 5))
print("MAE:",round(mean_absolute_error(Y,predictions), 5))
print("MAPE:", round(mean_absolute_percentage_error(Y,predictions), 5))
print("Theil:",round(Theil(Y,predictions), 5))
print(DAT)
DAT.plot(figsize=(19, 4))

```

```

plt.style.use('fivethirtyeight')
plt.show()
pyplot.style.use('fivethirtyeight')
def parser(x):
    return datetime.strptime(x, '%Y%m')
from pandas.plotting import autocorrelation_plot
series = read_csv('Credit.csv', header=0, parse_dates=[0], index_col=0,
squeeze=True)
X = series.values
Credit = pd.Series(series['LiquidityExp'])
Credit.plot()
pyplot.show()
print(Credit.head(77))
itog = Credit.describe()
print(Credit.describe())
row = [u'JB', u'p-value', u'skew', u'kurtosis']
jb_test = sm.stats.stattools.jarque_bera(Credit)
a = np.vstack([jb_test])
itog = SimpleTable(a, row)
print (itog)
test1 = sm.tsa.adfuller(Credit)
print ('adf: ', test1[0])
print ('p-value: ', test1[1])
print('Critical values: ', test1[4])
if test1[0]> test1[4]['5%']:
    print ('є одиничні корені - ряд не стаціонарний')
else:
    print ('одиничних коренів немає - ряд стаціонарний')

```

```

Creditdiff = Credit.diff(periods=1).dropna()
print(Creditdiff)
test = sm.tsa.adfuller(Creditdiff)
print ('adf: ', test[0])
print ('p-value: ', test[1])
print('Critical values: ', test[4])
if test[0]> test[4]['5%']:
    print ('є одиничні корені - ряд не стаціонарний')
else:
    print ('одиничних коренів немає - ряд стаціонарний')
m = Creditdiff.index[int(len(Creditdiff.index)/2+1)]
r1 = sm.stats.DescrStatsW(Creditdiff[m:])
r2 = sm.stats.DescrStatsW(Creditdiff[:m])
print ('p-value: ', sm.stats.CompareMeans(r1,r2).ttest_ind()[1])
Creditdiff.plot(figsize=(12,6))
pyplot.show()
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig = sm.graphics.tsa.plot_acf(Credit.values.squeeze(), lags=25, ax=ax1)
ax2 = fig.add_subplot(212)
fig = sm.graphics.tsa.plot_pacf(Credit, lags=25, ax=ax2)
def Summa(arr):
    add = 0
    for i in arr:
        add += i
    return add
def Theil(y_test,y_pred):

```

```

    return (np.sqrt(np.mean((np.array(test) -
np.array(predictions))**2)))/((np.sqrt(np.mean(np.array(test)**2)))+(np.sqrt(np.me
an(np.array(predictions)**2))))
def mean_absolute_percentage_error(y_test, y_pred):
    return np.mean(np.abs((np.array(test) - np.array(predictions)) / np.array(test))) *
100
history = [x for x in Credit]
history1 = [x for x in Credit]
for t in range(8):
    model = ARIMA(history, order=(2,1,0))
    model_fit = model.fit(dis=0)
    output = model_fit.forecast()
    history.append(output[0])
    print(output[0])
print(model_fit.summary())
residuals = DataFrame(model_fit.resid)
residuals.plot()
pyplot.show()
#residuals.plot(kind='kde')
pyplot.show()
print(residuals)
it=residuals.describe()
print(residuals.describe())
idx = pd.date_range('2001-04-01', periods=len(history), freq='3M')
idx1 = pd.date_range('2001-04-01', periods=len(history1), freq='3M')
ts = pd.Series(history, index=idx)
ts1 = pd.Series(history1, index=idx1)
pyplot.plot(ts, color='tomato')
pyplot.plot(ts1, color="dodgerblue")

```

```

pyplot.show()
Credit1 = pd.Series(series['LiquidityExp'])
size = int(len(Credit1)*0.7)
train, test = Credit1[0:size], Credit1[size:len(Credit1)]
hist = [x for x in train]
test = [y for y in test]
predictions = list()
for t in range(len(test)):
    model = ARIMA(hist, order=(1,1,0))
    model_fit = model.fit(dis=0, transparams=False)
    output = model_fit.forecast()
    yhat = output[0]
    predictions.append(yhat)
    obs = test[t]
    hist.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
print(model_fit.summary())
error = mean_squared_error(test, predictions)
DW = durbin_watson([test1 - predictions1 for test1, predictions1 in zip(test,
predictions)])
SS = np.sum((np.array(test) - np.array(predictions))**2)
print("Test MSE: %.3f" % error)
print("R-square:",round(abs(r2_score(test, predictions)), 5))
print("Durbin-Watson: %.3f" % DW)
print("Sum-Square: %.3f" % SS)
print("MSE:",round(mean_squared_error(test, predictions), 5))
print("MAE:",round(mean_absolute_error(test, predictions), 5))
print("MAPE:", round(mean_absolute_percentage_error(test, predictions), 5))
print("Theil:",round(Theil(test, predictions), 5))

```

```
pyplot.plot(test)
pyplot.plot(predictions, color='red')
pyplot.show()
```